Theoretical Computer Science 446 (2012) 75–97

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Approximating Markov processes through filtration[☆]

Chunlai Zhou [a,*], Mingsheng Ying [b,c]

[a] Department of Computer Science and Technology, School of Information, Renmin University of China, Beijing 100072, China

[b] State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

[c] Center of Quantum Computation and Intelligent Systems, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 15 June 2010
Received in revised form 27 January 2012
Accepted 15 March 2012
Communicated by B.P.F. Jacobs

Keywords:
Markov process
Probability logic
Coalgebra
Filtration

ABSTRACT

In this paper, we define a probabilistic version of filtration and use it to provide a finite approximation of Markov processes. In order to measure the approximation, we employ probability logic to construct the final Markov process and define a metric on the set of Markov processes through this logic. Moreover, we show that the set endowed with this metric is a Polish space. Finally we point to some questions connecting approximation to uniformity and approximate bisimilarity as topics for future research.

1. Introduction

Markov processes with continuous state spaces are important mathematical models in different physical sciences such as physics, biology, finance and computer sciences. With the ever-growing computer technology, we need to develop a theory of computational grip of this kind of important structures. If one is interested in computing them, one has to first discretize them. In order to achieve this goal, we must build a machinery to approximate Markov processes with continuous state spaces by discrete ones and also make sure that the approximating processes *preserve* all the essential properties especially the dynamic aspects of the original processes. The dynamics of the process is governed by the present state rather than by the past history of the process.

How does one know that the discretized process is a *faithful* approximation of the underlying continuous one? In other words, we should provide a notion of equivalence between continuous systems and discrete ones. Probabilistic bisimulation is central to study this kind of equivalence. Intuitively, two probabilistic bisimilar processes match transition probabilities for the same moves. Not only does probabilistic bisimulation enjoy some fundamental mathematical properties, most notably its characterization as a fixed point, but also it has a substantial *logical* meaning. Two processes are *logically equivalent* if they satisfy the same set of formulas whose syntax is specified as follows:

$$\varphi := \top \mid \varphi \wedge \varphi \mid \neg\varphi \mid L_r\varphi (r \in [0, 1] \cap \mathbb{Q})$$

where $\mathbb{Q}$ is the field of rationals. The formula $L_r\varphi$ says that the probability that the event φ will happen is at least r. One of the most important results for probabilistic bisimulation is the following Hennessy–Milner property.

Two processes are probabilistic bisimilar iff they are logically equivalent.

☆ The research of the first author is partly supported by NSF of China (Grant No: 60905036), Key project for Basic Research from the Ministry of Science and Technology of China (Grant No: 2012CB316205) and the German Academic Exchange Service (DAAD CodeNumber: A/08/94918); the second author is partly supported by NSF of China (Grant No: 60736011).

* Corresponding author. Tel.: +86 10 62510042.
E-mail addresses: czhou@ruc.edu.cn, chunlai.zhou@gmail.com (C. Zhou), yingms@mail.tsinghua.edu.cn (M. Ying).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2012.03.026

This property holds both for all discrete Markov processes [14] and for Markov processes with continuous state space with natural topological assumptions such as being Polish and analytical [14,20]. So the logic with the above syntax is also called a *Hennessy–Milner logic* for probabilistic bisimilarity. We can simply say that this logic *characterizes* probabilistic bisimilarity.

The Hennessy–Milner *property* can also be discretized. In process algebra [35], n-bisimilarity is defined and is shown to correspond to logical equivalence up to depth n. It is generalized to analyze transition systems with continuous state space in *qualitative* modal logic [8].[1] In some sense, this correspondence is a *discretized* form of the Hennessy–Milner property. Similarly this discretized form also holds for quantitative modal logics or simply probability logics. But we need to consider one more and the crucial factor in the discretization procedure: *quantitative indices* in probability formulas, which also makes our reasoning in probability logic much more involved than in qualitative modal logic. The natural method to incorporate this factor is to restrict all indices of the formulas that we consider to a finite set of rationals which are all multiples of 1/q for some natural number q.

In parallel, there is another closely-related discretization called *filtration*. In qualitative modal logic , given a Kripke model M and a finite set Γ of modal formulas, a filtration identifies states that satisfy the same subset of formulas in Γ and collapses the original model to a finite state one which preserves the satisfiability of formulas *in* Γ. Filtration for probability logics is quite similar. We only need to consider one more factor: probability indices and adapt the same approach as above for discretized Hennessy–Milner property.

The above-described discretizations are the main motivation of our approximation of Markov processes through filtration. The main contributions of the present paper about filtration are

- If a formula is satisfiable in a Markov process, it is also satisfied in a finite Markov process in the sense that it has only finitely many states and all its transition probabilities are multiples of 1/q for some natural number q.
- Given all finite approximants of a Markov process, we can reconstruct the dynamic aspects of the original one.

In order to capture precisely the approximation, we need a metric[2] to measure the difference between any two different Markov processes. Our metric is based on the smallest formula that distinguishes them. If the formula is very complicated, then a long sequence of observations are needed to tear them apart. So we believe that these two processes are very far away from each other. This metric takes two factors into account. The first one is the difference of transition probabilities of events. The second one is a discount factor c. This c will give more weight to the probability differences that arise earlier in the evolution of the process.

In the literature, Desharnais et al. [16] took a similar approach to ours. But they used *functional expressions* from [34] instead of probability formulas as building blocks to define a metric. So their machinery seems quite different from the Hennessy–Milner logic for probabilistic bisimilarity. In contrast, our approach to define the metric on Markov processes through *formulas* is in keeping with this logic. More importantly, we can easily show that the space of Markov processes with our metric is a Polish space.

van Breugel and Worrell [40] also defined a metric on probabilistic transition systems through *the final coalgebra of a functor* based on a metric on the space of Borel probability measures on a metric space. The existence of the final coalgebra requires a metric space because of Banach's fixed point theorem. They showed that Desharnais's metric defined by functional expressions and their metric through the final coalgebra are equivalent.

Our definition of metric is *purely logical* and will combine these two approaches into a uniform framework through probability logic. Not only are probability formulas employed to define a metric on Markov processes, but also they are used to construct the *final Markov process*. Our main contributions in this aspect are:

- a new definition of metric through probability formulas which also takes into account the discount factor;
- The space of Markov processes with this metric is shown to be a Polish space.

The paper is organized as follows. Section 2 will provide the background in Markov processes and probability logic, and construct the final Markov process out of maximally consistent set of formulas. We will elaborate filtration and approximation in Section 3. In Section 4, we define a new metric and show that the space of Markov processes endowed with this metric is a Polish space. The last section will point to some questions that we are going to address in the near future.

2. Markov processes and probability logic

In this section, we will present Markov processes in the coalgebraic setting and employ probability logic to construct a *final* Markov process.

[1] In this paper, qualitative modal logic refers to those with possible world semantics whereas probabilistic modal logic or simply probability logic stands for those with probabilistic transition semantics.

[2] Metrics in this paper are actually pseudometrics.

2.1. Preliminaries on measurable spaces and category theory

Let $\mathcal{A}$ be a (Boolean) algebra on a set X, i.e. a non-empty collection of subsets of X closed under complements and binary unions. $\mathcal{A}$ is a σ-algebra if it is also closed under countable unions. *A π-system* is a class of subsets of X closed under the formation of finite intersections. If $\mathcal{A}$ is a σ-algebra, then $\mathcal{X} = \langle X, \mathcal{A}\rangle$ is a *measurable space* and the elements of $\mathcal{A}$ are usually called *events* or *measurable subsets of* X. We write $\sigma(\mathcal{A}_0)$ for the smallest σ-algebra containing a given set $\mathcal{A}_0$ of subsets of $\mathcal{A}$. When $\sigma(\mathcal{A}_0) = \mathcal{A}$, we usually say that $\mathcal{A}_0$ *generates* $\mathcal{A}$. A *measurable function* $f : \langle X, \mathcal{A}\rangle \to \langle X', \mathcal{A}'\rangle$ is a function $f : X \to X'$ such that, for any $A' \in \mathcal{A}'$, $f^{-1}(A') \in \mathcal{A}$ where $\langle X', \mathcal{A}'\rangle$ is also a measurable space. A set function $\mu : \mathcal{A} \to [0,\infty]$ on $\mathcal{A}$ in X is *finitely additive* if $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ whenever A_1 and A_2 are disjoint elements of $\mathcal{A}$. μ is called a (countably additive) *measure* if it satisfies the following conditions:

1. $\mu(\emptyset) = 0$;
2. $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ where $\{A_i\}_{i=1}^{\infty}$ is a pairwise disjoint sequence of events of $\mathcal{A}$.

The second property is usually called the *countable additivity*. The measure μ is *finite* or *infinite* as $\mu(X) < \infty$ or $\mu(X) = \infty$. If $\mu(X) \leq 1$, then μ is called a *subprobability measure*. If $\mu(X) = 1$, then μ is called a *probability measure*. A metric space $\langle X, \rho\rangle$ is *complete* if any Cauchy sequence has a limit in X, and ρ is called a *complete* metric. A topological space $\langle X, \tau\rangle$ is called *separable* if it has a countable dense subset. A *Polish space* $\langle X, \tau\rangle$ is a separable topological space which is metrizable through a complete metric. The *Borel σ-algebra* $\mathcal{B}(X, \tau)$ for the topology τ is the smallest σ-algebra that contains τ. An *analytical* space is the image of a Polish space under a continuous function from one Polish space to another. The following three theorems are useful for our further results. The interested reader can find their proofs and more details in [7,28].

Theorem 2.1. *Let $\mathcal{A}$ be an algebra,*

1. *Any measure μ on $\mathcal{A}$ is continuous from above, meaning that if $\{A_n : n \in N\} \subseteq \mathcal{A}$ is a non-increasing sequence whose intersection belongs to $\mathcal{A}$, with at least one A_n having finite measure, then $\mu(\bigcap_{n\in N} A_n) = \lim_{n\to\infty} \mu(A_n)$.*
2. *Let $\mu : \mathcal{A} \to [0,\infty]$ be finitely additive with $\mu(X) < \infty$. Then μ is a measure if it is continuous at $\emptyset$, i.e., $\lim_{n\to\infty} \mu(A_n) = 0$ for any non-increasing sequence $\{A_n : n \in \mathbf{N}\}$ with $\bigcap_n A_n = \emptyset$.*

Theorem 2.2. *Suppose that μ_1 and μ_2 are finite measures on $\sigma(\mathcal{A})$, where $\mathcal{A}$ is a π-system and $\sigma(\mathcal{A})$ is the σ-algebra generated by $\mathcal{A}$. If μ_1 and μ_2 agree on $\mathcal{A}$, then they agree on $\sigma(\mathcal{A})$.*

Theorem 2.3. *Let $f : (X, \mathcal{A}_X) \to (Y, \mathcal{A}_Y)$ be a function between measurable spaces, and $\mathcal{C}$ generate $\mathcal{A}_Y$. Then f is measurable iff $f^{-1}(C) \in \mathcal{A}_X$ for each $C \in \mathcal{C}$.*

For any measurable space $\mathcal{X}$, we obtain a measurable space $\Delta(\mathcal{X})$ of all probability measures on $\mathcal{X}$ with the σ-algebra $\mathcal{A}_\Delta$ generated by the sets $\{\beta^r(A) : r \in [0,1], A \in \mathcal{A}\}$, where $\beta^r(A) = \{\mu \in \Delta(\mathcal{X}) : \mu(A) \geq r\}$. If $f : \langle X, \mathcal{A}\rangle \to \langle X', \mathcal{A}'\rangle$ is measurable, $\Delta(f)$ is defined to be a function from $\Delta(\mathcal{X})$ to $\Delta(\mathcal{X}')$ with the following property:

$$\text{for any } \mu \in \Delta(\mathcal{X}), A' \in \mathcal{A}', \Delta(f)(\mu)(A') = \mu(f^{-1}(A')).$$

Theorem 2.4 (*Giry [25]*)**.** *$\Delta(f)$ is measurable.*

This theorem is the essential technical measure-theoretical result in this paper. The category **Meas** has the measurable spaces as objects and the measurable functions as morphisms, with the usual functional composition of morphisms.

Let **C** and **D** be categories. A *functor* F from **C** to **D** is a mapping that

- associates to each object $X \in \mathbf{C}$ an object $F(X) \in \mathbf{D}$,
- associates to each morphism $f : X \to Y \in \mathbf{C}$ a morphism $F(f) : F(X) \to F(Y) \in \mathbf{D}$ such that the following two conditions hold:
  1. $F(id_X) = id_{F(X)}$ for every object $X \in \mathbf{C}$;
  2. $F(g \circ f) = F(g) \circ F(f)$ for all morphisms $f : X \to Y$ and $g : Y \to Z$.

That is, functors must preserve identity morphisms and composition of morphisms.

A functor from a category to itself is called an *endofunctor*. The above defined Δ is an *endofunctor* on **Meas** and is actually the well-known Giry functor [25].

The functor $\mathbf{C} \to \mathbf{D}$ which maps every object of **C** to a fixed object X in **D** and every morphism in **C** to the identity morphism on X is called a *constant* or *selection* functor. The *identity functor* on X is the functor that maps every object and morphism to themselves. The Cartesian product $X_1 \times X_2$ of two sets has associated projections $\pi_j : X_1 \times X_2 \to X_j$ for $j \in \{1, 2\}$. The coproduct $X_1 + X_2$ of X_1 and X_2 is their disjoint union, with injective insertion function $in_j : X_j \to X_1 + X_2$ for $j \in \{1, 2\}$. Each element of $X_1 + X_2$ is equal to $in_j(x)$ for a unique j and a unique $x \in X_j$. Sometimes we use *inl* and *inr* for in_1 and in_2, respectively. Also these constructions lift to measurable spaces. The σ-algebra of the product space $X_1 \times X_2$ is generated by the products $A_1 \times A_2$ of measurable sets A_j from each factor X_j, or equivalently by the inverse image $\pi_j^{-1}(A_j)$ of the measurable sets from each factor. The σ-algebra of the co-product space $X_1 + X_2$ is generated by the insertion $in_j(A_j)$ of the measurable sets A_j from each summand X_j. Note that all the associated projections, insertions are all measurable functions.

Two measurable functions $f_1 : X_1 \to X_1'$ and $f_2 : X_2 \to X_2'$ have a measurable product $f_1 \times f_2 : X_1 \times X_2 \to X_1' \times X_2'$ and a measurable co-product $f_1 + f_2 : X_1 + X_2 \to X_1' + X_2'$, where

- $(f_1 \times f_2)(x_1, x_2) = (f_1(x_1), f_2(x_2))$;
- $(f_1 + f_2)(in_j(x_j)) = in_j(f_j(x_j))$.

The *product* $T_1 \times T_2$ of two functors on **Meas** is the functor that acts on the space by $X \to T_1X \times T_2X$, and on morphisms by $f \mapsto T_1(f) \times T_2(f)$. The *co-product* functor $T_1 + T_2$ has $X \to T_1X + T_2X$ and $f \mapsto T_1(f) + T_2(f)$. $\mathbf{1}$ is the terminal object functor. This functor maps each object to the terminal object $\mathbf{1}$ in **Meas** which is the singleton space. The interested reader may consult [2] for the basics about category theory and [7] for measure theory.

The class of *measurable polynomial functors* is the smallest class of functors on **Meas** containing the identity functor *Id*, the constant functor M for each measurable space M, and closed in the following ways: if U and V are measurable polynomials, then so are $U + V$, $U \times V$ and ΔU. Our main results in this paper concerns the functor $\Delta(Id + \mathbf{1})$.

Let $T : \mathcal{C} \to \mathcal{C}$ be an endofunctor on the category $\mathcal{C}$; then a *coalgebra* for T is a pair (A, f) consisting of an object $A \in \mathcal{C}$ and a T-morphism $f : A \to TA$. A *coalgebra morphism* h from a coalgebra (X, f) to another coalgebra (X', f') is a T-morphism X to X' such that $f' \circ h = T(h) \circ f$, that is to say that the following diagram commutes:

$$\begin{tikzcd} X \arrow[r, "h"] \arrow[d, "f"'] & X' \arrow[d, "f'"] \\ T(X) \arrow[r, "T(h)"'] & T(X') \end{tikzcd}$$

In particular, each Δ-morphism is a measurable function. A coalgebra (A, f) is *final* if for every coalgebra (B, g) there is exactly one coalgebra morphism from (B, g) to (A, f). In theoretical computer science, the existence of final coalgebras is usually closely related to the semantics of recursive types [32]. In game theoretical economics, the existence of final coalgebras also proved the equivalence of implicit and explicit descriptions of belief types [30]. In some sense, we try to unify in this paper the perspectives from such different fields as game theoretical economics, coalgebra and Markov processes into a uniform framework through probability logic. The interested reader may refer to [37] for the theory of coalgebras and to [4] for game theoretical economics.

2.2. Markov processes and their logical characterization

Definition 2.5. A *transition (sub)probability function* T on a measurable space $\mathcal{X} = \langle X, \mathcal{A} \rangle$ is a function from $X \times \mathcal{A}$ to $[0, 1]$ satisfying the following two conditions:

- for each $x \in X$, $T(x, \cdot)$ is a (sub)probability measure, and
- for each $A \in \mathcal{A}$, $T(\cdot, A)$ is a measurable function. $\lhd$

Definition 2.6. A *Markov process* M is a structure $\langle X, i, \mathcal{A}, T \rangle$, where $\langle X, \mathcal{A} \rangle$ is measurable space, i is the initial state[3] in X, and T is a subprobability transition function. $\lhd$

We will not consider the initial state until Section 4. In this case, we simply write M as $\langle X, \mathcal{A}, T \rangle$.

Proposition 2.7. *Markov processes are coalgebras of the sub-probability measure functor* $S(X) = \Delta(X + \mathbf{1})$.

Let t be a measurable function from $\mathcal{X} = \langle X, \mathcal{A} \rangle$ to $\Delta(\mathcal{X}) = \langle \Delta(X), \mathcal{A}_\Delta \rangle$. Define $T_t : X \times \mathcal{A} \to [0, 1]$ as $T_t(x, A) = t(x)(A)$. It is easy to check that T_t is a transition probability function. Conversely, if T is a transition probability function, the function $t_T : X \to \Delta(X)$ defined below is measurable [44,45]:

$$t_T(x)(A) := T(x, A) \quad \text{for each } x \in X, A \in \mathcal{A}.$$

So a Markov process $M = \langle X, \mathcal{A}, T \rangle$ with a transition probability function T is a *coalgebra* (X, t_T) of Δ. In game theoretical economics, a transition probability function is called a *type* function [30]. Generally, a Markov process $M = \langle X, \mathcal{A}, T \rangle$ is the coalgebra (X, t_T) of the functor $S(X) = \Delta(X + \mathbf{1})$ where, for each $x \in X$ and $A \in \mathcal{A}$, $t_T(x)(inl(A)) = T(x)(A)$.

Our following definitions of zigzag morphism (from [14]) and of a final Markov process just paraphrase the above notions of coalgebra morphism and of a final coalgebra, respectively.

Definition 2.8. A function $f : \langle X, \mathcal{A}, T \rangle \to \langle X', \mathcal{A}', T' \rangle$ is a *zigzag morphism* if it is measurable, and the following equality holds:

$$T(x, f^{-1}(A')) = T'(f(x), A'), \quad \text{for any } x \in X, A' \in \mathcal{A}'. \quad \lhd$$

3 In the definition of metric on Markov processes in Section 4, we need the initial states.

Definition 2.9. A Markov process $M = \langle X, \mathcal{A}, T\rangle$ is *final* if, for every Markov process $M' = \langle X', \mathcal{A}', T'\rangle$, there is a unique zigzag morphism from M' to M. $\lhd$

Definition 2.10. The two Markov processes $\langle X, \mathcal{A}, T\rangle$ and $\langle X, \mathcal{A}', T'\rangle$ are probabilistically *bisimilar* or simply bisimilar if there is a Markov process $\langle X'', \mathcal{A}'', T''\rangle$ with two *surjective* zigzag morphisms $h' : X'' \to X'$, and $h : X'' \to X$. $\lhd$

Let $\langle X, f\rangle$ and $\langle X', f'\rangle$ be two Markov processes, considered as coalgebras. They are probabilistically bisimilar (or simply bisimilar) if there exists a third coalgebra $\langle X'', f''\rangle$ with two surjective coalgebra morphisms h and h' such that the following diagram commutes:

$$\begin{tikzcd}
X \arrow[d, "f"'] & X'' \arrow[l, "h"'] \arrow[r, "h'"] \arrow[d, "f''"] & X' \arrow[d, "f'"] \\
\Delta(X+\mathbf{1}) & \Delta(X''+\mathbf{1}) \arrow[l, "\Delta(h+\mathbf{1})"] \arrow[r, "\Delta(h'+\mathbf{1})"'] & \Delta(X'+\mathbf{1})
\end{tikzcd}$$

One important result about Markov processes is that there is a Hennessy–Milner logic to characterize the above probabilistic bisimulation. A *formula* φ of the logic is formed by the following syntax:

$$\varphi := \top \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid L_r\varphi (r \in Q \cap [0, 1])$$

where Q is the field of rationals. $\mathcal{L}_0$ denotes the language of this simple syntax and $\mathcal{L}_0^+$ is $\mathcal{L}_0$ without negation. It is well-known [14] that $\mathcal{L}_0^+$ is rich enough to provide the Hennessy–Milner property for probabilistic bisimulation. However, since a deductive system for probability logic plays an important role in this paper, we choose to keep negation to get a standard completeness result.

Lemma 2.11. *For a fixed Markov process $M = \langle S, \mathcal{A}, T\rangle$, there exists a unique satisfaction relation $\models$ between the state w of M and modal formulas φ in $\mathcal{L}_0$ that satisfies the following clauses; moreover, the associated interpretation $[[\varphi]]_M = \{w \in S : M, w \models \varphi\}$ is a measurable set, for all formulas φ.*

- $M, w \models \top$ *for all* $w \in S$;
- $M, w \models \varphi_1 \wedge \varphi_2$ *iff* $M, w \models \varphi_1$ *and* $M, w \models \varphi_2$;
- $M, w \models \neg\varphi$ *iff* $M, w \not\models \varphi$;
- $M, w \models L_r\varphi$ *iff* $T(w)([[\varphi]]_M) \geq r$, *where* $[[\varphi]]_M := \{w \in S : M, w \models \varphi\}$.

Proof. The two statements in the lemma can be proved simultaneously, by mutual recursion on the complexity of formulas φ. The crucial step is based on the fact that the definition of T guarantees that $[[\varphi]]_M \in \mathcal{A}$ for all formulas φ, especially, if $[[\varphi]]_M \in \mathcal{A}$, then $[[L_r\varphi]]_M = (T(\cdot, [[\varphi]]_M))^{-1}([r, 1])$, which is measurable according to the definition of T. $\square$

A formula φ is *valid* in the Markov process M if $M \models \varphi$, i.e. for all states $w \in S$, $M, w \models \varphi$. It is *valid in a class of Markov processes* $\mathcal{C}$ if, for each $M \in \mathcal{C}$, $M \models \varphi$. Let Γ be a set of formulas and φ a formula. Γ is satisfied at a state w of M if all formulas in Γ are satisfied at w, which is denoted $M, w \models \Gamma$. And $\Gamma \models \varphi$ means

for any $w \in M$ in any Markov process M, $M, w \models \Gamma$ implies $M, w \models \varphi$.

A *satisfied* theory at a state x of a Markov process is the set of all formulas in $\mathcal{L}_0$ that are satisfied at this state. In the following sections, we will construct a final Markov process by these satisfied theories. First let Ω denote the set of all satisfied theories. The description map d from a Markov process $M' = \langle \Omega', \mathcal{A}', T'\rangle$ to Ω is defined as follows: for each $w' \in \Omega'$, $d(w') = \{\varphi \in \mathcal{L}_0 : M', w' \models \varphi\}$. Note that, since the association of d to M' is inessential, we simply omit M' from the supposed subscript. The following 2 lemmas are needed to show Theorem 2.17. We put the proofs of these two lemmas in the Appendix.

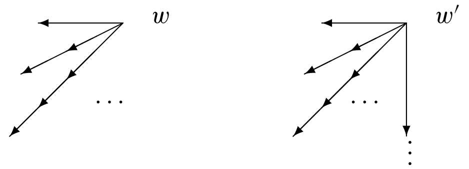
Lemma 2.12. *Let $M_1 = \langle \Omega_1, \mathcal{A}_1, T_1\rangle$ and $M_2 = \langle \Omega_2, \mathcal{A}_2, T_2\rangle$ be Markov processes. If $f : M_1 \to M_2$ is a zigzag morphism, then, for any formula φ, $f^{-1}([[\varphi]]_{M_2}) = [[\varphi]]_{M_1}$, where $[[\varphi]]_{M_1} = \{w \in \Omega_1 : M_1, w \models \varphi\}$ and $[[\varphi]]_{M_2} = \{w \in \Omega_2 : M_2, w \models \varphi\}$.*

Lemma 2.13. *Let $M_1 = \langle \Omega_1, \mathcal{A}_1, T_1\rangle$ and $M_2 = \langle \Omega_2, \mathcal{A}_2, T_2\rangle$ be Markov processes. If $f : M_1 \to M_2$ is a zigzag morphism, then, f preserves the satisfiability of formulas, i.e., for any $w \in \Omega_1$, $d(w_1) = d(f(w_1))$, where d is the above description map.*

The following is the well-known theorem about the Hennessy–Milner property or expressivity of Markov processes [14,22].

Theorem 2.14. *Let $\langle \Omega_1, \mathcal{A}_1, T_1\rangle$ and $\langle \Omega_2, \mathcal{A}_2, T_2\rangle$ be two Markov processes whose Ω_i's ($i = 1, 2$) are Polish spaces and $\mathcal{A}_i$'s ($i = 1, 2$) are Borel σ-algebra. They are bisimilar iff they satisfy the same set of formulas of $\mathcal{L}_0$ or $\mathcal{L}_0^+$.*

Remark 2.15. In qualitative modal logic [8], bisimilarity implies modal equivalence. In other words, if two states are bisimilar, then they satisfy the same set of formulas. Although the converse holds for finite state systems [31], it does not for infinite state ones. The eminent counterexample is as follows:



It is easy to check that w and w' satisfies the same set of modal formulas but they are not bisimilar [8].

The above theorem is just the counterpart of the Hennessy–Milner theorem in qualitative modal logic. The connection between syntactic equivalence and probabilistic bisimulation is explored in detail in the monograph [23]. In order to secure the Hennessy–Milner property, we must impose some topological conditions on Markov processes such as above being Polish or being analytical [14]. However, there is *behavioral equivalence* studied in [11,33,18,38] for which one need not impose anything.

2.3. Probability logic and final coalgebra

In this section, we will develop a deductive system to characterize the semantic consequence relation $\models$. Our deductive system is not about theoremhood but about the deducibility relation $\Gamma \vdash \varphi$, from sets of formulas to formulas, what is intended to capture the idea that φ is deducible from members of Γ with aid of various axioms and rules of inference. The definition of $\vdash$ is to be syntactic, depending only on the symbolic pattern of formulas and basic set-theoretic properties of sets of them. We attempt to show that $\vdash$ is identical to the above semantically defined consequence relation $\models$, thereby characterizing $\models$ *proof-theoretically*. Without further notice, all rationals below are between 0 and 1 inclusively.

Probability Logic
Axioms:

- (A0) propositional calculus
- (A1) $L_0\varphi$
- (A2) $\neg L_r\bot$, for $0 < r \leq 1$
- (A3) $L_r(\varphi \wedge \psi) \wedge L_t(\varphi \wedge \neg\psi) \to L_{r+t}\varphi$, for $r + t \leq 1$
- (A4) $\neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \to \neg L_{r+s}\varphi$, for $r + s \leq 1$
- (A5) $L_r\varphi \to \neg L_s\neg\varphi$, for $r + s > 1$

Rules:

- (AR)(Assumption Rule) If $\varphi \in \Gamma$ or φ is an instance of axioms, then $\Gamma \vdash \varphi$;
- (CR) (Cut Rule) If $\Gamma \vdash \psi$ for all $\psi \in \Sigma$ and $\Sigma \vdash \varphi$, then $\Gamma \vdash \varphi$,
- (DR) Deduction Rule: $\Gamma \cup \{\varphi\} \vdash \psi$ implies $\Gamma \vdash \varphi \to \psi$.
- (ARCH): $\{\gamma \to L_s\varphi : s < r\} \vdash \gamma \to L_r\varphi$.
- (CAR) (Countable Additivity Rule) $\Gamma \vdash \varphi$ implies $\{L_p\psi : \psi \in \bigwedge_\omega \Gamma\} \vdash L_p\varphi$ where Γ is countable and $\bigwedge_\omega \Gamma$ is the set of conjunctions of finite subsets of Γ.

Since we reason not about probability measures but about general subprobability measures, we do not include $L_1\top$ in the above list as Goldblatt did in [27]. The last rule (CAR) characterizes exactly the continuity from above property of operators L_r. It is needed in the proof of countable additivity of the measures defined on the canonical models [27]. Note that the rules (CR) and (DR) are needed but not essential in our axiomatization. Their only function is to aid the deductions through other inference rules. A relation $\vdash \subseteq 2^{\mathcal{L}_0} \times \mathcal{L}_0$ is a *probability logic* if the following rules hold: modus ponens, AR, ARCH, CR, DR, CAR and uniform substitution (that is, if $(\Gamma, \varphi) \in \vdash$, then so do all of its substitution instances). Observe that the rule (ARCH) is the only rule that is really about the indices of the modalities. Since the index set $Q \cap [0, 1]$ has the *Archimedean property*, i.e., the property of having no infinitely small elements, the fact that the rule has infinitely many premises seems unavoidable. It is easy to check that the consequence relation $\models$ defined in last section is a probability logic. The system with (A0–A5) plus (AR), (CR) and (DR) was originally proposed by Aumann [3]. But this system was proved to be incomplete [29]. The rule (ARCH) was added in [44] to show the resulting system is weakly complete. Goldblatt [27] further added (CAR) to obtain a *strongly complete* system, which is the above deductive system about $\vdash$. Aumann constructed the canonical probability model *without* any deductive system [3]. Following the same lines as in [3], Heifetz and Mongin [29] proved the existence of the universal type space, which is essentially the final coalgebra for $\Delta(M \times Id)$ for some fixed space M. From the perspective of coalgebra, Moss and Viglizzo [36] further showed that every measurable polynomial functor has a final coalgebra. Goldblatt [27] provided deduction systems for those coalgebras.

Although much of this part is adapted from [27,36,45], our focus is not on the deduction system itself but on defining a pseudometric to measure approximation on the collection of Markov processes through a final coalgebra $M_{\mathcal{L}_0}$ constructed out of maximally consistent sets of formulas in probability logic. In [36], Moss and Viglizzo showed that a coalgebra $c^* : Id^* \to T(Id^*)$ is final, where each element of Id^* is the set of formulas *with sort Id* (each sort corresponds to an ingredient of T) that are satisfied at some state of some coalgebra of T. In Appendix B, we define a translation $^\circ$ from the language $\mathcal{L}(T)$ with sorts to the language $\mathcal{L}_0$ and show that $^\circ$ is actually a coalgebra isomorphism between c^* and our $M_{\mathcal{L}_0}$. This translation is of independent interest.

Given a probability logic Σ_p, $(\Gamma, \varphi) \in \Sigma_p$ is denoted as $\Gamma \vdash_{\Sigma_p} \varphi$. A set Γ of formulas is Σ_p*-consistent* if $\Gamma \nvdash_{\Sigma_p} \perp$. A probability logic Σ_p is *Lindenbaum* if every Σ_p-consistent set of formulas has a maximally Σ_p-consistent extension. *Probability logic* Σ_s is the least Lindenbaum probability logic. In order to streamline our presentation and to emphasize our main results in the following sections, we will put the proof of the completeness of probability logic in Appendix A. It is easy to see that the following distribution rule is derivable from (CAR) and (DR):

(DIS) If $\vdash \varphi \leftrightarrow \psi$, then $\vdash L_r\varphi \leftrightarrow L_r\psi$.

The canonical Markov process $M_{\mathcal{L}_0}$ is defined as follows:

- $\Omega_{\mathcal{L}_0} = \{\Gamma \subseteq \mathcal{L}_0 : \Gamma$ is a maximally Σ_s-consistent set of formulas$\}$.
- $[\varphi] = \{w \in \Omega_{\mathcal{L}_0} : \varphi \in w\}$;
- $\mathcal{A}_{\mathcal{L}_0}$ is the σ-algebra generated by $\{[\varphi] : \varphi \in \mathcal{L}_0\}$;
- The canonical measurable space is defined as $\langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0} \rangle$;
- define $T_{\mathcal{L}_0}(w)([\varphi]) = sup\{r \in [0, 1] \cap \mathbb{Q} : L_r\varphi \in w\}$;

The above defined $T_{\mathcal{L}_0}(w)$ on $\{[\varphi] : \varphi \in \mathcal{L}_0\}$ uniquely determines the subprobability measure $T_{\mathcal{L}_0}(w)$ on the σ-algebra $\mathcal{A}_{\mathcal{L}_0}$. This unique extension is guaranteed by Theorem 2.2. Such a defined canonical model $M_{\mathcal{L}_0} = \langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0}, T_{\mathcal{L}_0} \rangle$ is a Markov process.

Theorem 2.16 (*Strong Completeness*). *For any $\Gamma \subseteq \Omega_{\mathcal{L}_0}$ and $s \in \Omega_{\mathcal{L}_0}$,*

$$\Gamma \vdash_{\Sigma_s} \varphi \quad \text{if and only if} \quad \Gamma \models \varphi.$$

Theorem 2.17. *Given a Markov process M, the description map d is a zigzag morphism and moreover the unique zigzag morphism from M to $M_{\mathcal{L}_0}$. This means that $M_{\mathcal{L}_0}$ is the final Markov process.*

3. Filtration

This section will discuss the main topic: approximation through filtration. We do need such a strong logic as Σ_s to construct the final probabilistic transition system. However, for filtration, a much weaker probability logic Σ_+ will suffice (for more details about Σ_+, one may refer to [45]):

Probability Logic Σ_+

- (A0) propositional calculus
- (A1) $L_0\varphi$
- (A2) $\neg L_r\perp$, for $0 < r \leq 1$
- (A3) $L_r(\varphi \wedge \psi) \wedge L_t(\varphi \wedge \neg\psi) \to L_{r+t}\varphi$, for $r + t \leq 1$
- (A4) $\neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \to \neg L_{r+s}\varphi$, for $r + s \leq 1$
- (A5) $L_r\varphi \to \neg L_s\neg\varphi$, for $r + s > 1$
- (DIS) If $\vdash \varphi \leftrightarrow \psi$, $\vdash L_r\varphi \leftrightarrow L_r\psi$.
- (ARCH): If $\vdash \gamma \to L_s\varphi$ for all $s < r$, then $\vdash \gamma \to L_r\varphi$.

Probability logic Σ_+ is the smallest set of formulas that contains all propositional tautologies in $\mathcal{L}_0$ and (A1–A5), and is closed under modus ponens, DIS, ARCH and uniform substitution (that is, if φ belongs to Σ_+, then so do all of its substitution instances). It is clear that Σ_+ is much simpler than Σ_s. The main difference between Σ_s and Σ_+ is that Σ_s is concerned about the deducibility relation while Σ_+ is about the theoremhood. However, it can be shown [46] that a formula φ is a theorem in Σ_+ iff it is also a theorem in Σ_s, i.e. $\emptyset \vdash_{\Sigma_s} \varphi$. Since we deal with only finite consistency in the section on filtration, the simpler Σ_+ will suffice. For this reason, finitely Σ_s-consistent or finitely Σ_+-consistent will be simply called finitely consistent.

Lemma 3.1 (*Lindenbaum Lemma*). *If Ξ is a finitely-consistent set of formulas, then there is a maximally finitely-consistent Ξ' such that $\Xi \subseteq \Xi'$.*

Definition 3.2. The *depth* $dp(\varphi)$ of a formula φ is defined inductively:

- $dp(\top) := 0$;
- $dp(\neg\varphi) := dp(\varphi)$;
- $dp(\varphi_1 \wedge \varphi_2) := max\{dp(\varphi_1), dp(\varphi_2)\}$;
- $dp(L_r\varphi) := dp(\varphi) + 1$. $\lhd$

Now we define a local language $\mathcal{L}_0(q, d)$ to be the largest set of formulas satisfying the following conditions:

- The indices of formulas in $\mathcal{L}_0(q, d)$ are multiples of $1/q$;
- The formulas in $\mathcal{L}_0(q, d)$ are of depth $\leq d$;
- Logically equivalent formulas are regarded the same.

The above q is called the *accuracy* of the language $\mathcal{L}_0(q, d)$. In particular, $\mathcal{L}_0[\psi]$ is defined as $\mathcal{L}_0(q_\psi, d_\psi)$ where q_ψ is the accuracy of ψ i.e., the least common multiple of all denominators of the indices appearing in ψ and d_ψ the depth of ψ. $I[\psi]$ is the finite set of all rationals in the form of $p/q_\psi \in [0, 1]$; and it is called the *index set* of the language $\mathcal{L}_0[\psi]$. Note that $\mathcal{L}_0[\psi]$ is finite. In general, let $\mathcal{L}_0(q, d)$ be the set of formulas having accuracy q and depth at most d modulo logical equivalence where two formulas θ_1 and θ_2 are *logically equivalent* if $\theta_1 \leftrightarrow \theta_2$ is provable in the deductive system Σ_+. Each formula in $\mathcal{L}_0(q, d)$ is logically equivalent to a finite disjunction consisting only of non-equivalent disjuncts, each of the disjuncts being a conjunction consisting only of non-equivalent conjuncts, each conjunct being either itself in $\mathcal{L}_0(q, d-1)$ or being obtainable from some formula in $\mathcal{L}_0(q, d-1)$ by prefixing it either with a modality L_r or $\neg L_r$ where r is a multiple of $1/q$. By induction on the depth d, we can show that $\mathcal{L}_0(q, d)$ is finite and hence $\mathcal{L}_0[\psi]$ is finite. In the following, we will not distinguish between the equivalence classes in $\mathcal{L}_0(q, d)$ and their representatives.

Definition 3.3. Assume that Θ is a finite set of formulas. Let q_Θ be the accuracy of Θ, i.e., the least common denominator of all rationals occurring in Θ and d_Θ the largest depth of formulas in Θ. Θ is *maximal consistent in the language* $\mathcal{L}(q_\Theta, d_\Theta)$ if it is consistent and no subset of $\mathcal{L}(q_\Theta, d_\Theta)$ properly containing Θ is consistent. $\lhd$

Lemma 3.4 (*Lindenbaum Lemma for Finite set of Formulas*). *Assume that Θ is a finite set of formulas. If $d_\Theta \leq d$ and q is a multiple of q_Θ for some integers q and d, then there is a maximal consistent extension $\Theta(q, d)$ in the language $\mathcal{L}_0(q, d)$ such that $\Theta \subseteq \Theta(q, d)$.*

In the following, a language is meant to be the whole language $\mathcal{L}_0$ or the finite language $\mathcal{L}_0(q, d)$ for some integers q and d.

Given a Markov process $M = \langle \Omega, \mathcal{A}, T\rangle$, two states s and s' in Ω are *logically equivalent up to* $\mathcal{L}_0(q, d)$ (denoted as $s \backsim_{(q,d)} s'$) if they satisfy the same set of formulas in $\mathcal{L}_0(q, d)$. In other words,

$$s \backsim_{(q,d)} s' \text{ if, for all formulas } \varphi \in \mathcal{L}_0(q, d), M, s \models \varphi \Leftrightarrow M, s' \models \varphi.$$

It is easy to see that the above defined $\backsim_{(q,d)}$ is an equivalence relation. We denote the equivalence class of a state s with respect to $\backsim_{(q,d)}$ by $|s|_{(q,d)}$. Note that $|s|_{(q,d)}$ is $\mathcal{A}$-measurable. The mapping $s \mapsto |s|_{(q,d)}$ from a state to its equivalence class is called the *natural map*. Recall that $I_q = \{p/q : 0 \leq p/q \leq 1, p \text{ is a natural number}\}$. Any Markov process $M_{(q,d)} := \langle \Omega_{(q,d)}, \mathcal{A}_{(q,d)}, T_{(q,d)}\rangle$ with $T_{(q,d)}(|s|_{(q,d)})$ a subprobability measure on $\mathcal{A}_{(q,d)}$ for each $|s|_{(q,d)}$ satisfying the following conditions is called *a filtered Markov process* of M:

1. $\Omega_{(q,d)} = \{|s|_{(q,d)} : s \in \Omega\}$;
2. $\mathcal{A}_{(q,d)} = 2^{\Omega_{(q,d)}}$, i.e. the powerset of the filtered carrier set;
3. for all formulas $L_r\varphi \in \mathcal{L}_{(q,d)}$ where $r \in I_q$, $T_{(q,d)}(|s|_{(q,d)}, [\varphi]_M) \geq r \Leftrightarrow T(s, [[\varphi]]_M) \geq r$ where $[\varphi]_M = \{|s|_{(q,d)} \in \Omega_{(q,d)} : M, s \models \varphi\}$.

Note that, in order to distinguish filtered Markov models from canonical models in the following Remark 3.7, we put the index tuples as subscripts.

Remark 3.5. It is easy to prove that $\Omega_{(q,d)}$ is finite and so is $\mathcal{A}_{(q,d)}$. The third clause is the most important in the definition, and it will be used to show the following filtration theorem. The two directions of the clause correspond exactly to the two similar characteristic clauses in the definition of a filtered model in qualitative modal logic (page 78 in [8]). It is easy to see that the following inequality holds:

$$|T_{(q,d)}(|s|_{(q,d)}, A') - T(s, \textstyle\bigcup_{|s|_{(q,d)} \in A'} |s|_{(q,d)})| \leq 1/q.$$

It would be interesting to compare the third clause to zigzag morphism. In zigzag morphism, the two related transition probabilities are equal to each other while, in the third clause of the above definition, the related probabilities do not need to be equal, but are required to satisfy the same finite set of inequalities. It is easy to check that the following $T^s_{(q,d)}$ satisfies the third clause:

for each $|s|_{(q,d)} \in \Omega_{(q,d)}$, $T^s_{(q,d)}(|s|_{(q,d)}) = T(s')$ for some $s' \in |s|_{(q,d)}$ in the sense that

$$T^s_{(q,d)}(|s|_{(q,d)}, A') = T(s', \textstyle\bigcup_{|s|_{(q,d)} \in A'} |s|_{(q,d)}) \text{ for all } A' \in \mathcal{A}_{(q,d)}.$$

So such defined $T^s_{(q,d)}$ induces a natural filtered model $\langle \Omega_{(q,d)}, \mathcal{A}_{(q,d)}, T^s_{(q,d)}\rangle$.

Lemma 3.6. *Let $[\varphi]_M$ denote the set $\{|s|_{(q,d)} : s \in [[\varphi]]_M\}$ for $\varphi \in \mathcal{L}_0(q, d)$. The following holds: $2^{\Omega_{(q,d)}} = \{[\varphi]_M : \varphi \in \mathcal{L}_0(q, d)\}$.*

Proof. We only need propositional reasoning to show this lemma. It is easy to see that $\{[\varphi]_M : \varphi \in \mathcal{L}_0(q,d)\} \subseteq 2^{\Omega(q,d)}$. Given any set $\Xi \in 2^{\Omega(q,d)}$, define $\bar{\Xi} = \{\bar{s}_{(q,d)} : |s|_{(q,d)} \in \Xi\}$ where $\bar{s}_{(q,d)} = \{\varphi \in \mathcal{L}_0(q,d) : M, s' \models \varphi \text{ for all } s' \in |s|_{(q,d)}\}$.

Let φ_Ξ be the disjunction of the conjunctions of formulas in each $\bar{s}_{(q,d)}$ in $\bar{\Xi}$. φ_Ξ is logically equivalent to a formula φ'_Ξ in $\mathcal{L}_0(q,d)$. It is straightforward to check that $\Xi = [\varphi'_\Xi]_M$. So we have shown the other direction. □

Remark 3.7. It is important to note the connection of the above filtered Markov process to the finite canonical model that was used to prove the completeness of different deductive systems for probability logics [29,44,45], which is also the inspiration of our work in this paper. A canonical model $M(q,d) := \langle \Omega(q,d), 2^{\Omega(q,d)}, T(q,d)\rangle$ is constructed as follows:

- $\Omega(q,d)$ is the set of all maximally consistent subsets of formulas in $\mathcal{L}(q,d)$;
- The powerset $2^{\Omega(q,d)}$ is the set of events;
- for each $\Gamma \in \Omega(q,d)$, $T(q,d)(\Gamma)$ is defined either through the Rockafellar Lemma [29] or through a maximally consistent extension to the language $\mathcal{L}_0$ [45].

In [45], the first author has shown that not only does this canonical model have finitely many states, but also it is actually *computable*, the transition probabilities are computable. In some sense, the above filtered Markov process $M_{(q,d)}$ can be regarded as a kind of submodel of the canonical model $M(q,d)$ in the sense that each state $|s|_{(q,d)}$ can be represented as $\bar{s}$, which is a state in $\Omega(q,d)$ and $M_{(q,d)}$ can be embedded into the canonical model $M(q,d)$ by $|s| \mapsto \bar{s}$ with a variation $1/q$ in transition probabilities.

Example 3.8. Each approximation through filtration is *guided* by a local language $\mathcal{L}_0(q,d)$. Here we adapt the example on page 179 in [15] so that we may compare the approximation there with our following one through filtration.

The state space is $S := \{s,t\} \cup [0,3]$, which is a continuous state space. The transition probabilities on the space are defined as follows:

- If $x \in [0,1]$,
$$T(x,[0,y)) = \frac{x+y}{4}, \quad \text{where } 0 \le y \le 1;$$
$$T(x,\{1\}) = \frac{1-x}{4}$$
$$T(x,(1,1+y]) = \frac{y}{4}$$
$$T(x,(2,2+y]) = \frac{xy}{4}$$
- if $x \in (1,2]$, $T(x,\{s\}) = 1$;
- if $x \in (2,3]$, $T(x,\{t\}) = \frac{1}{8}$.

In order to make our approximation transparent and simple, we drop the labels in the example in [15], change the last transition from 1 to $\frac{1}{8}$ and consider the language $\mathcal{L}_0$. First we deal with the filtration through the local language $\mathcal{L}_0(2,1)$. Note that $\mathcal{L}_0(2,1)$ does not contain any propositional letters. For example, $L_{\frac{1}{2}}\top$ is a formula in this set. According to the above transition function T, we know that

- if $x \in [0,1]$, $T(x,S) = \frac{3+x}{4}$;
- if $x \in (1,2]$, $T(x,S) = 1$;
- if $x \in (2,3]$, $T(x,S) = \frac{1}{8}$

hence the equivalence classes with respect to the local language $\mathcal{L}_0(2,1)$ are: $[0,1)$, $[1,2]$, $(2,3]$ and $\{s,t\}$. Note that the only significant change to the partition of the state space in the definition of T is the move of 1 from $[0,1]$ to $[1,2]$ because $T(1,S) = 1$ and also $T(x,S) = 1$ for all $x \in (1,2]$. Now we define the transition probabilities on the filtered state space $S_{(2,1)} := \{[0,1),[1,2],(2,3],\{s,t\}\}$ according to the $T^s_{(q,d)}$ defined in Remark 3.5.

1. If $x' = [0,1)$, define $T_{(2,1)}(x) = T(0)$, namely,
   - $T_{(2,1)}(x',[0,1)) = T(0,[0,1)) = \frac{1}{4}$,
   - $T_{(2,1)}(x',[1,2]) = T(0,[1,2]) = \frac{1}{2}$,
   - $T_{(2,1)}(x',(2,3]) = T(0,(2,3]) = 0$;
   - $T_{(2,1)}(x',\{s,t\}) = T(0,\{s,t\}) = 0$.
2. If $x' = [1,2]$, define $T_{(2,1)}(x) = T(1)$, namely,
   - $T_{(2,1)}(x',[0,1)) = T(1,[0,1)) = \frac{1}{4}$,
   - $T_{(2,1)}(x',[1,2]) = T(1,[1,2]) = \frac{1}{4}$,
   - $T_{(2,1)}(x',(2,3]) = T(1,(2,3]) = \frac{1}{4}$;
   - $T_{(2,1)}(x',\{s,t\}) = T(1,\{s,t\}) = 0$.

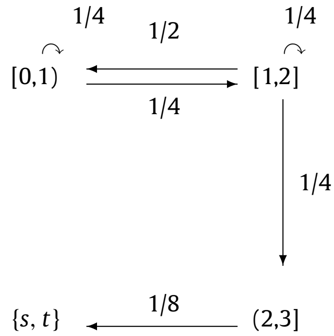


Fig. 1. The filtered Markov process $M_{(2,1)}$.

3. If $x' = (2, 3]$, define $T_{(2,1)}(x) = T(3)$, namely,
   - $T_{(2,1)}(x', [0, 1)) = T(3, [0, 1)) = 0$,
   - $T_{(2,1)}(x', [1, 2]) = T(3, [1, 2]) = 0$,
   - $T_{(2,1)}(x', (2, 3]) = T(3, (2, 3]) = 0$.
   - $T_{(2,1)}(x', \{s, t\}) = T(3, \{s, t\}) = \frac{1}{8}$.

   Note that in each equality the second arguments are treated differently. For example, $[0, 1)$ in $T_{(2,1)}(x, [0, 1))$ is an equivalence class with respect to $\mathcal{L}_0(2, 1)$ while $[0, 1)$ in $T(0, [0, 1))$ is a subset of S.

We draw the filtered Markov process $M_{(2,1)}$ in an informal way in Fig. 1.

Now we are constructing the filtration with respect to the local language $\mathcal{L}_0(2, 2)$. Here we only need to consider the transition probabilities from each $\sim_{(2,1)}$-equivalence class to all finite unions of these equivalence classes. Essentially we obtain a table accommodating all necessary information about these transition probabilities:

	[0,1)	[1,2]	(2,3]	[0,2]	[1, 3]	[0,1) ∪ (2,3]	[0,3]	$\{s, t\}$
[0,1)	$\frac{x+1}{4}$	$\frac{2-x}{4}$	$\frac{x}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{2x+1}{4}$	$\frac{3+x}{4}$	0
{1}	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	0
(1,2]	0	0	0	0	0	0	0	1
(2,3]	0	0	0	0	0	0	0	$\frac{1}{8}$
$\{s, t\}$	0	0	0	0	0	0	0	0

For example, the entry $\frac{2x+1}{4}$ indicates the transition probability from $[0, 1)$ to the union $[0, 1) \cup (2, 3]$. If $x \in [0, \frac{1}{2})$, then $\frac{2x+1}{4} \in (0, \frac{1}{2})$; if $x = 1/2$, then $\frac{2x+1}{4} = \frac{1}{2}$; if $x \in (\frac{1}{2}, 1)$, then $\frac{2x+1}{4} \in (\frac{1}{2}, 1)$. This implies that, for $[0, 1)$, the language $\mathcal{L}_0(2, 2)$ distinguishes among $[0, \frac{1}{2})$, $\{\frac{1}{2}\}$ and $(\frac{1}{2}, 1)$. After applying the same analysis to other entries (actually mainly to those in the first row), we can obtain all $\sim_{(2,2)}$-equivalence classes as follows:

$$S_{(2,2)} := \left\{\{0\}, \left(0, \frac{1}{2}\right), \left\{\frac{1}{2}\right\}, \left(\frac{1}{2}, 1\right), \{1\}, (1, 2], (2, 3], \{s, t\}\right\}.$$

Here we take the $\sim_{(2,2)}$-equivalence class $(0, \frac{1}{2})$ as an illustration how to define a filtered model. For $(0, \frac{1}{2})$, we define $T_{(2,2)}((0, \frac{1}{2})) = T(\frac{1}{4})$ in the sense that, for each $\sim_{(2,2)}$-equivalence class C, $T_{(2,2)}((0, \frac{1}{2}))(C) = T(\frac{1}{4})(C)$. For example, $T_{(2,2)}((0, \frac{1}{2}))((0, \frac{1}{2})) = T(\frac{1}{4})((0, \frac{1}{2})) = \frac{3}{16}$.

3.1. *Finite approximation through filtration*

Compared to the *analytical* approximation in [15], our approximation through filtration is *logical* in the sense that this kind of approximation is *guided* by finite logical languages $\mathcal{L}_0(q, d)$. In the remainder of this section, we use it to show the main results in [15] in a simpler way.

Theorem 3.9 (*Filtration Theorem*). *Let $M := \langle \Omega, \mathcal{A}, T\rangle$ be a Markov process and $M_{(q,d)} := \langle \Omega_{(q,d)}, \mathcal{A}_{(q,d)}, T_{(q,d)}\rangle$ be a filtered Markov process through $\mathcal{L}_0(q, d)$. For any formula $\varphi \in \mathcal{L}_0(q, d)$ and any point $s \in \Omega$,*

$$M, s \models \varphi \text{ iff } M_{(q,d)}, |s|_{(q,d)} \models \varphi.$$

Equivalently, $[[\varphi]]_M = \bigcup_{M_{(q,d)}, |s'|_{(q,d)} \models \varphi} |s'|_{(q,d)}$.

Proof. We prove this by induction on the formula φ.

1. Boolean case. This is straightforward from the fact that $|s|_{(q,d)}$ is a maximal and consistent set of formulas in the above local language.
2. Crucial case: $\varphi = L_r\varphi'$ where r is a multiple of $1/q$.

$$
\begin{aligned}
M_{(q,d)}, |s|_{(q,d)} \models L_r\varphi' &\Leftrightarrow T_{(q,d)}(|s|_{(q,d)})([[\varphi']]_{M_{(q,d)}}) \geq r \\
&\Leftrightarrow T_{(q,d)}(|s|_{(q,d)})(\{|s'|_{(q,d)} : M_{(q,d)}, |s'|_{(q,d)} \models \varphi'\}) \geq r \\
&\Leftrightarrow T(s)\left(\bigcup_{M_{(q,d)}, |s'|_{(q,d)} \models \varphi'} |s'|_{(q,d)}\right) \geq r \\
&\Leftrightarrow T(s)([[\varphi']]_M) \geq r \\
&\Leftrightarrow M, s \models L_r\varphi'.
\end{aligned}
$$

The fourth equivalence is based on the induction hypothesis and the third is on the third clause of the definition of filtration. □

Corollary 3.10. *Let $M := \langle \Omega, \mathcal{A}, T\rangle$ be a Markov process and $M_{(q,d)} := \langle \Omega_{(q,d)}, \mathcal{A}_{(q,d)}, T_{(q,d)}\rangle$ be a filtered Markov process through $\mathcal{L}_0(q, d)$. For any formula $\varphi \in \mathcal{L}_0(q, d)$ and any point $s \in \Omega$,*

$$|T_{(q,d)}(|s|_{(q,d)})([[\varphi]]_{M_{(q,d)}}) - T(s)([[\varphi]]_M) \leq 1/q.$$

Remark 3.11. The above Filtration Theorem says that any formula of the finite restricted type which is satisfiable in a Markov process is also satisfied in a *finite state* filtered Markov process. Although there are infinitely many subprobability measures for the filtered process, they satisfy the same *finite* set of linear inequalities as the corresponding transition subprobabilities in the original process, which is indicated in the third clause of our definition of filtration. According to the Finite Model Theorem(Theorem 3.13) in [45], we know that each transition subprobability can be a multiple of $1/n$ for some natural number n. This also means that we can make all transition subprobabilities in each filtered Markov process to be rational and still the above filtration theorem holds. Compared with the measure-theoretical approach to finite model property in [15], our method is much simpler and in keeping with filtration in qualitative modal logic [8]. In the remainder of this section, we will use filtration to show a similar result to Theorem 4.5 in [15].

Now we show that one can reconstruct the original process M from the approximants $M_{(n,n)}$ where $M_{(n,n)}$ is a filtered process of the original M through a finite set $\mathcal{L}_0(n, n)$ of formulas. Just as Theorem 4.5 in [15], we recover mainly all the transition subprobability information, i.e., the dynamic aspects of the process. For any two states x and x' of M, $x \approx x'$ denotes that they satisfy the same set of formulas in $\mathcal{L}_0$. It is easy to see that $\approx$ is an equivalence relation.

Theorem 3.12. *Assume that $M = \langle S, \mathcal{A}, T\rangle$ is a Markov process that is maximally collapsed in the sense that $M = M/\approx$. If we are given all finite state approximants $M_{(n,n)}$, then we can recover M.*

Proof. We can recover the state space by just taking the union of states at any level of any approximants $M_{(n,n)}$ because, according to the Filtration Theorem,

$$[[\varphi]]_M = \bigcup_{M_{(n,n)}, |s|_{(n,n)} \models \varphi} |s|_{(n,n)}.$$

Since M is a maximally collapsed Markov process, $\mathcal{A}$ is generated by $\mathcal{A}_{\mathcal{L}_0} := \{[[\varphi]]_M : \varphi$ is a formula in the language of $\mathcal{L}_0\}$(by the Corollary 4.15 of [13]). It is easy to check that $\mathcal{A}_{\mathcal{L}_0}$ is a field. For any $s \in S$ and any formula φ (in the language $\mathcal{L}_0(n, n)$ for $n > k$ for some k), define

$$\mu(s)([[\varphi]]_M) = \lim_{n\to\infty} T_{(n,n)}(|s|_{(n,n)})([[\varphi]]_{M_{(n,n)}}).$$

Claim 1. $\lim_{n\to\infty} T_{(n,n)}(|s|_{(n,n)})([[\varphi]]_{M_{(n,n)}})$ *exists.*

In order to show the claim, it suffices to show that, for any ε, there is an N such that, if $n \geq N$, $|T_{(n,n)}(|s|_{(n,n)})([\varphi]_{M_{(n,n)}}) - T_{(n+p,n+p)}(|s|_{(n+p,n+p)})([\varphi]_{M_{(n+p,n+p)}})| \leq \varepsilon$ for any natural number p. For a given ε, there is an N such that $1/N < \varepsilon/2$. Also for any natural numbers $p, n > N$,

$$
\begin{aligned}
&|T_{(n,n)}(|s|_{(n,n)})([[\varphi]]_{M_{(n,n)}}) - T_{(n+p,n+p)}(|s|_{(n+p,n+p)})([[\varphi]]_{M_{(n+p,n+p)}})| \\
&\quad \leq |T_{(n,n)}(|s|_{(n,n)})([[\varphi]]_{M_{(n,n)}}) - T(s)([[\varphi]]_M)| + |T_{(n+p,n+p)}(|s|_{(n+p,n+p)})([[\varphi]]_{M_{(n+p,n+p)}}) - T(s)([[\varphi]]_M)| \\
&\quad \leq 1/n + 1/(n+p) \\
&\quad \leq 2/N \\
&\quad \leq \varepsilon.
\end{aligned}
$$

The second equality follows from the third clause of the definition of filtration and from the filtration theorem. By Cauchy's criterion for sequence convergence [28], we know that the above limit exists. This is to say, $\mu(s)([[\varphi]]_M)$ is well-defined.

Claim 2. *For any $s \in S$ and for any formula φ, $\mu(s)([[\varphi]]_M) = T(s)([[\varphi]]_M)$.*

Note that

$$\begin{aligned} 0 &\leq |\mu(s)([[\varphi]]_M) - T(s)([[\varphi]]_M)| \\ &= lim_{n\to\infty}|T(s)([[\varphi]]_M) - T_{(n,n)}(|s|_{(n,n)})([[\varphi]]_{M_{(n,n)}})| \\ &\leq lim_{n\to\infty} 1/n \\ &= 0. \end{aligned}$$

That is to say, $\mu(s)([[\varphi]]_M) = T(s)([[\varphi]]_M)$. The third step similarly follows from the third clause of the definition of filtration and the second from Corollary 3.10. So we have shown that, for any $s \in S$, the above defined $\mu(s)$ and $T(s)$ matches on the algebra $\mathcal{A}_{\mathcal{L}_0}$. This implies that the subprobability measure $\mu^*(s)$ on the σ-algebra $\mathcal{A}$ generated by $\mu(s)$ is the same as $T(s)$. So we have recovered the original Markov process. In other words, the Markov process M is uniquely determined by its finite approximants $M_{(n,n)}$. □

Just like the Dedekind's cut for reals, we want to use finite rational Markov processes (in the sense that all transition subprobabilities are rationals) to approximate any Markov process. In Remark 3.11, we know that all transition subprobabilities in each filtered process can be rational. So any Markov process M can be recovered by all its finite approximants $M_{(n,n)}$ with all transition subprobabilities rational.

Corollary 3.13. *Assume that $M = \langle S, \mathcal{A}, T\rangle$ is maximally collapsed in the sense that $M = M/\approx$. Then M is uniquely determined by its rational approximants $M_{(n,n)}$ with all transition subprobabilities rational.*

This corollary is similar to Theorem 4.9 in [15]. From this Corollary, we know that any Markov process can be approximated by rational finite filtered processes, which provides a countable basis for Markov processes. In the next section, we will make this approximation precise by defining a metric on the collection of Markov processes.

Since the collection of finite rational Markov processes is countable and any Markov process can be approximated by finite rational processes in a certain metric space, it is reasonable to expect that the collection of Markov processes endowed with this metric is a separable space. In the following section, we will simulate the work in [16] to show that the collection of Markov processes is also a Polish space. In some sense, we answer the question raised in Section 6 of [15] whether the class of Markov processes can be a Polish space.

4. Metric

Our main motivation for the following metrics for Markov processes comes from two sources. The first is the technical definition of such metrics through a class of functions expounded by Kozen [34] to generalize logic to handle probabilistic phenomena [16]. We know from the logical characterization of bisimulation in [14] that if two processes are not bisimilar, there will be a formula that distinguishes them. We want the metric to formalize the intuition that two processes are bisimilar iff the "distance" between them is zero. Our definition of the metric is based on the *smallest* formula that distinguishes them. Also we need to consider the fact that the process might differ immediately but do so with probabilities that are very close. In this case, we should think that they might be very "far" away from each other. Just as in [16], we introduce a discount factor c to give more weight to the probability difference that arises earlier in the evolution of the process.

So our metrics take into account the following two factors: the complexity of the distinguishing formula which is measured by its depth and the *maximal* probability difference by which any formula distinguishes them.

Our second source is the definition of metric through a final coalgebra in van Breugel and Worrell [40]. Their metric is based on finding a final coalgebra for a certain functor on the category of metric spaces and non-expansive maps. This final coalgebra is also a metric space and thus naturally gives a metric on the state spaces of any Markov processes through the unique map induced by finality.

So we combine these two to provide a (pseudo)metric for Markov processes. First we construct a final coalgebra for the functor S which maps X to $\Delta(X + \mathbf{1})$ through probability logic Σ_s in the restricted language $\mathcal{L}_0$, which is defined on Section 2.2. Second we define a metric on this final coalgebra.

Recall that $\Omega_{\mathcal{L}_0}$ is the set of all maximally Σ_s-consistent set of formulas in the language $\mathcal{L}_0$. For any formula φ in the language $\mathcal{L}_0$, $[\varphi]$ denotes the subset $\{s \in \mathcal{L}_0 : \varphi \in s\}$. It is easy to check that the subset $\mathcal{A}^0_{\mathcal{L}_0} = \{[\varphi] : \varphi$ is a formula in the language $\mathcal{L}_0\}$ of the powerset $2^{\Omega_{\mathcal{L}_0}}$ is an algebra. Let $\mathcal{A}_{\mathcal{L}_0}$ denote the σ-algebra generated by the algebra $\mathcal{A}^0_{\mathcal{L}_0}$. Define the transition subprobability function on the canonical model as follows: for any $s \in \Omega_{\mathcal{L}_0}$,

$$T_{\mathcal{L}_0}(s)([\varphi]) = sup\{r \in Q \cap [0, 1] : L_r\varphi \in s\}$$

Theorem 2.17 has shown that $M_{\mathcal{L}_0} := \langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0}, T_{\mathcal{L}_0}\rangle$ is the final Markov process for the above functor S.

For any $s, s' \in \Omega_0^C$, define $d^c_{\mathcal{L}_0}(s, s') = \sup\{c^{sdp(\varphi)} \cdot \mid T_{\mathcal{L}_0}(s)([\varphi]_0) - T_{\mathcal{L}_0}(s')([\varphi]_0) \mid : \varphi \in \mathcal{L}_0\}$ where $sdp(\varphi)$ denotes the smallest *depth* of the formulas which are logically equivalent to φ in Σ_s[4] and $c(\in [0, 1])$ is a discount factor. It is easy to check that it is a (pseudo)metric. Note that being smallest depth guarantees that the metric is independent of the choice of representative formulas. It follows from the definition of $d^c_{\mathcal{L}_0}$ that

Proposition 4.1. *For any two states $s, s' \in \Omega_{\mathcal{L}_0}$, the following are equivalent:*

- *they are bisimilar;*
- *they satisfy the same set of formulas;*
- $d^c_{\mathcal{L}_0}(s, s') = 0$.

The following is the main theorem of this section which tells us that the final Markov processes with the metric $d^c_{\mathcal{L}_0}$ is a Polish space. For simplicity, we will show only the case when $c = 1$. In this case, we simply write $d^1_{\mathcal{L}_0}$ as $d_{\mathcal{L}_0}$. Actually the proofs of other cases when $c < 1$ are similar.

Theorem 4.2. *$(M_0^C, d_{\mathcal{L}_0})$ is a complete metric space.*

Proof. Assume that $\{s_n\}_{n=1}^{\infty}$ is a Cauchy sequence of states of $\Omega_{\mathcal{L}_0}$, i.e. given an $\varepsilon > 0$, there is a natural number N_0 such that, if $n, m \geq N_0$, $d_{\mathcal{L}_0}(s_n, s_m) < \varepsilon$, namely
$sup\{\mid T_{\mathcal{L}_0}(s_n)([\varphi]_0) - T_{\mathcal{L}_0}(s_m)([\varphi]_0) \mid : \varphi \in \mathcal{L}_0\} < \varepsilon/4$.

Fix this ε in this proof. It is easy to see that, for each $\varphi \in \mathcal{L}_0$, $\{T_{\mathcal{L}_0}(s_n)([\varphi])\}_{n=1}^{\infty}$ is a Cauchy sequence in the real interval $[0, 1]$. This also implies that $\lim_{n\to\infty} T_{\mathcal{L}_0}(s_n)([\varphi])$ exists.

Set $\rho : \mathcal{A}_{\mathcal{L}_0} \to [0, 1]$ as follows:

$$\rho([\varphi]) = \lim_{n\to\infty} T_{\mathcal{L}_0}(s_n)([\varphi]).$$

Claim 3. *ρ is a finitely additive set function on $\mathcal{A}_{\mathcal{L}_0}$.*

Proof of this claim: Assume that $[\varphi_i](1 \leq i \leq m)$ are disjoint.

$$\begin{aligned}\rho\left(\bigcup_{i=1}^m [\varphi_i]\right) &= \lim_{n\to\infty} T_{\mathcal{L}_0}(s_n)\left(\bigcup_{i=1}^m ([\varphi_i])\right)\\ &= \lim_{n\to\infty} \sum_{i=1}^m T_{\mathcal{L}_0}(s_n)([\varphi_i])\\ &= \sum_{i=1}^m \lim_{n\to\infty} T_{\mathcal{L}_0}(s_n)([\varphi_i])\\ &= \sum_{i=1}^m \rho(([\varphi_i])).\end{aligned}$$

So indeed ρ is a finitely additive set function.

Claim 4. $\lim_{n\to\infty} d(T_{\mathcal{L}_0}(s_n), \rho) = 0$ *where* $d(T_{\mathcal{L}_0}(s_n), \rho) = \sup\{|T_{\mathcal{L}_0}(s_n)([\varphi] - \rho([\varphi])| : \varphi \in \mathcal{L}_0\}$.

Proof of the claim: First note that for each φ, there is a natural number N_φ such that, if $n \geq N_\varphi$ then $\mid T_{\mathcal{L}_0}(s_n)([\varphi]) - \rho([\varphi]) \mid < \varepsilon/4$. Pick up a natural number N'_φ that is bigger than both N_φ and N_0. If $n \geq N_0$, then

$$\begin{aligned}|T_{\mathcal{L}_0}(s_n)([\varphi]) - \rho([\varphi])| &\leq |T_{\mathcal{L}_0}(s_n)([\varphi]) - T_{\mathcal{L}_0}(s_{N'_\varphi})([\varphi])| + |T_{\mathcal{L}_0}(s_{N'_\varphi})([\varphi]) - \rho([\varphi])|\\ &\leq \varepsilon/4 + \varepsilon/4\\ &= \varepsilon/2.\end{aligned}$$

The first ε came from the fact that $N'_\varphi \geq N_0$ and the second one from $\lim_{n\to\infty} T_{\mathcal{L}_0}(s_n)([\varphi]_0) = \rho([\varphi]_0)$. In short,

if $n \geq N_0$, then $|T_{\mathcal{L}_0}(s_n)([\varphi]_0) - \rho([\varphi]_0)| \leq \varepsilon/2$ for all $\varphi \in \mathcal{L}_0$.

It follows immediately that $\lim_{n\to\infty} d(T_0^c(s_n), \rho) = 0$.

Claim 5. *ρ is a (countably additive) subprobability measure on $\mathcal{A}_{\mathcal{L}_0}$.*

Proof of the claim: The essential part is to show that ρ is countably additive. Because of Theorem 2.1, it suffices to show the following statement:

If $[\varphi_1] \supseteq [\varphi_2] \supseteq \cdots$ such that $\bigcap_{i=1}^{\infty}[\varphi_i] = \emptyset$, then $\lim_{i\to\infty} \rho([\varphi_i]) = 0$.

---

[4] We have not yet found any algorithm to compute $sdp(\varphi)$ for each φ.

Since $T_{\mathcal{L}_0}(s_n)$ is a subprobability measure for each n, $\lim_{i\to\infty} T_{\mathcal{L}_0}(s_n)([\varphi_i]) = 0$. In particular, for the above natural number N_0, $\lim_{i\to\infty} T_{\mathcal{L}_0}(s_{N_0})([\varphi_i]) = 0$. That is to say, there is a natural number M_{N_0} such that, if $i \geq M_{N_0}$, then $T_{\mathcal{L}_0}(s_{N_0}([\varphi_i]) < \varepsilon/2$. We have that if $i \geq M_{N_0}$,

$$\begin{aligned} |\rho([\varphi_i])| &\leq |\rho([\varphi_i]) - T_0^c(s_{N_0})([\varphi_i])| + |T_{\mathcal{L}_0}(s_{N_0})([\varphi_i])| \\ &\leq \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

Note that M_{N_0} depends only on ε. So we have finished the proof of the claim.

Consider the set

$$\Gamma := \{L_r\varphi, \neg L_t\varphi : \varphi \in \Phi_{\mathcal{L}_0}, r, t \in \mathbb{Q}, r \leq \rho([\varphi]_0), t > \rho([\varphi]_0)\}.$$

By a similar argument to that of Theorem 2.16, we know that, since Γ is defined from ρ which is the canonical measure on the canonical measurable space $\langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0}\rangle$, it is Σ_s-consistent. So it is contained in a maximally Σ_s-consistent set of formulas $s_\infty \in M_{\mathcal{L}_0}$. It is easy to see that such a s_∞ is unique because any state in $M_{\mathcal{L}_0}$ is a maximally Σ_s-consistent set of formulas in $\mathcal{L}_0$. In other words, Γ has a unique maximally Σ_s-consistent extension in $\mathcal{L}_0$. We conclude that $\lim_{n\to\infty} d_{\mathcal{L}_0}(s_n, s_\infty) = 0$ and $(M_{\mathcal{L}_0}, d_{\mathcal{L}_0})$ is a complete metric space. □

Now we discuss how to define a metric on the collection of Markov processes *with initial states*. Let $M_1 = \langle \Omega_1, s_1, \mathcal{A}_1, T_1\rangle$ and $M_2 = \langle \Omega_2, s_2, \mathcal{A}_2, T_2\rangle$ be two Markov processes. Since $M_{\mathcal{L}_0} = \langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0}, T_{\mathcal{L}_0}\rangle$ is final, there is a unique map f_0^1 from M_1 to $M_{\mathcal{L}_0}$ induced by the finality and similarly a unique map f_0^2 from M_2 to $M_{\mathcal{L}_0}$. From Theorem 2.17, we know that these two maps are the description maps from M_1 and M_2 to $M_{\mathcal{L}_0}$, respectively. We simply define the metric between these two processes as follows:

$$d_{\mathcal{L}_0}(M_1, M_2) := d_{\mathcal{L}_0}(f_0^1(s_1), f_0^2(s_2)).$$

Corollary 4.3. *With respect to the above defined metric, the approximation in Theorem 3.13 can be also stated as* $\lim_{n\to\infty} d_{\mathcal{L}_0}((M_{(n,n)}, M) = 0$.

Theorem 4.4. *The metric space of Markov processes is separable.*

Proof. This theorem follows directly from Theorem 3.13 in the last section. □

Corollary 4.5. *The space of Markov processes with $d_{\mathcal{L}_0}$ is a Polish space.*

Remark 4.6. In the literature there are two well known approaches to define metrics on probabilistic transition systems or Markov processes. The first metric $d_{\mathcal{F}^c}$ in [16] is through a real-valued modal logic $\mathcal{F}^c$ (defined in the next paragraph) adapted from the work by Kozen [34]. The second one d_C is to provide a coalgebraic definition [40]. Van Breugel and Worrell [40] showed that these two approaches are equivalent in the sense that

$$\text{for all } x_1, x_2 \in M_f, \quad \frac{d'_{\mathcal{F}^c}(x_1, x_2)}{c} = d'_C(x_1, x_2)$$

where c is the discount factor and M_f is the final probabilistic transition system or the final Markov process in the setting of this paper.

Desharnais et al. [16] defined a set of functions which are sufficient to characterize bisimulation. They defined a set of functional expressions by giving an explicit syntax. For each $c \in (0, 1]$, we consider a family $\mathcal{F}^c$ of functional expressions generated by the following grammar:

$$f := \mathbf{1} \mid \mathbf{1} - f \mid \langle a\rangle f \mid min(f_1, f_2) \mid \sup_{i\in\mathbf{N}} f_i \mid f \ominus q$$

where q is a rational. Each collection $\mathcal{F}^c$ of functional expressions induces a distance function as follows:

$$d_{\mathcal{F}^c}(\mathcal{P}, \mathcal{Q}) = \sup_{f\in\mathcal{F}_c} |f_{\mathcal{P}}(p_0) - f_{\mathcal{Q}}(q_0)|$$

where p_0 and q_0 are initial states of two Markov processes $\mathcal{P}$ and $\mathcal{Q}$ respectively. One of the main theorem in [16] says that the distance of two processes is 0 iff they are bisimilar.

Van Breugel and Worrell [40] employed the machinery from category theory to show that there is a final Markov process and then define a metric through this final system. Moreover, they showed that their approach to define metric is the same as that by Desharnais et al. through functional expressions [40]. Their metric is closely related to the Hutchinson metric in the literature [24].

In contrast, our method is *purely logical* in the sense that it combines these two approaches in a uniform framework of probability logic Σ_s. Not only do we employ probability logic to construct final Markov process, but also we define a metric on the final one through the set of formulas instead through functional expressions and the metric for Markov processes is defined through the unique map to the final one induced by the finality.

Compared to the metric $d_{\mathcal{F}^c}$ defined through functional expressions by Desharnais et al. [16], our $d^c_{\mathcal{L}_0}$ is *more appropriate and direct* in the sense that the complexities of distinguishing *logical formulas* instead of distinguishing functional expressions are used to measure the distances among Markov processes. The fundamental result in [16] states that, given any formula $\varphi \in \mathcal{L}_0$ and $c \in (0, 1]$, there exists a functional expression $f_\varphi \in \mathcal{F}^c_+$ such that for every state s of any Markov process M, $f_\varphi(s) > 0$ iff $M, s \models \varphi$. So the value of any functional expressions in $\mathcal{F}^c$ corresponds to a quantitative measure of the extent to which the state *satisfies* a formula of $\mathcal{L}_0$. That is to say, given any two states s and s', the distance $d_{\mathcal{F}^c}(s, s')$ measures the difference of *satisfactions* of formulas in $\mathcal{L}_0$ at these two states. However, $d^c_{\mathcal{L}_0}$ provides a quantitative valuation of the combination of the difference of *transition probabilities* of different events (represented by *logical formulas*) and of the complexity of the distinguishing formulas (captured by the depths of these formulas). In this sense, $d^c_{\mathcal{L}_0}$ is more appropriate than $d_{\mathcal{F}^c_0}$ as a pseudometric to characterize the differences of the *dynamics* of Markov processes.

Although the approaches to define $d^c_{\mathcal{L}_0}$ and d_C are different, they are equivalent in measure-theoretical topology, namely, they induce the same metric topology in the setting of [40,39]. Here we take $c = 1$ as an example. Let $M^f = \langle S, t, d_S\rangle$ be the final coalgebra of the subprobability functor which maps X to $\Delta(X + \mathbf{1})$ where t is the transition function and d_S induces a Borel σ-algebra $\mathcal{B}_S$ on S. If $\mathcal{B}_S$ is also generated by $\{[[\varphi]]_{M^f} : \varphi \in \mathcal{L}_0\}$ where $[[\varphi]]_{M^f} = \{x \in S : M^f, x \models \varphi\}$, one may apply the well-known theorem on weak convergence in measure theory (Theorem 2.5.17 in [24], Theorem 25.8 in [7]) to show that, on the final Markov process M^f, the metric d_C (in the form of Hutchinson metric here) induces *the same metric topology* as that by $d_{\mathcal{L}_0}$.

One advantage *with this new metric* $d^c_{\mathcal{L}_0}$ is that we can show *in a natural analytical way* that the class of Markov processes is a Polish space.

5. Future work

We are developing a non-metric theory of approximating Markov transition systems through uniform topological spaces.[5] As far as we know, all the approaches of approximating Markov transition systems employ metrics to measure the distances among different systems [16,12,41] and all associated metric spaces there are *uniformly* complete. In some sense, qualitative reasoning is more essential than quantitative one. They are more interested in reasoning about closeness instead in the numerical value of the nearness of different systems. In general topology, this kind of reasoning is related to uniformity properties. To put it more precisely, uniform spaces provide a *uniform* framework for approximation of Markov transitive systems. This approach can also ease the technical arguments involving rationals and reals in the previous literature.

We note that there is a predecessor to this kind of work. In [19], Doberkat has observed that the complete metric space as a type model proposed by MacQueen, Plotkin and Sethi is a special case of complete uniformities. The main technical innovation of the paper is the formulation of uniformities on a set of types and the establishment of the existence of recursively defined types through a fixed-point theorem over uniform structures.

Our strategy is as follows. First we would define a functor on the category of uniform topologies as objects and uniform continuity as morphisms. Next we would consider a coalgebraic definition of morphisms and *uniformly* topological bisimilarity. The main result that we are looking for is an appropriate *Hennessy–Milner logic* characterizing uniform bisimilarity. The existence of the fixed point theorem for complete uniform spaces that Doberkat proved in [19] is expected to show this kind of Hennessy–Milner property for *continuous time* stochastic logic with a *fixed point* operator [22,21].

Alfaro et al. [1] introduced two fundamental notions of equivalences and metrics for two-player games over *finite* state space for an infinite number of rounds where the goals of the winning games are expressed in the quantitative μ-calculus. These two notions can be regarded as the canonical extensions to games, of the classical notions of metrics and bisimulation for probabilistic transition systems. We expect to explore the connection of this extension with continuous time stochastic logic with a fixed point operator [22].

We have in mind some other approximations of Markov processes in [10], especially approximation through averaging in [9]. In parallel to the approaches there, we are developing a *logical* deductive system for expectation and following the same line as in this paper to approximate Markov processes.

We also expect to apply the above coalgebraic perspective of complete uniform spaces to approximate bisimilarity defined by Ying [43,42]. In Ying's theory on approximate bisimilarity, the actual numerical values of the differences are less important than the qualitative reasoning as in uniform spaces. This similarity may provide a setting to combine the Hennessy–Milner logic for uniform spaces and approximate bisimilarity.

The probabilistic transition systems that we consider in this paper are discrete time, continuous state space models. Now we are generalizing our above results about this kind of models to *continuous-time Markov processes*(CTMPs), which are an important class of stochastic processes that have been widely used in software engineering [5,6]. Continuous-time stochastic logic provides a characterization of bisimilarity for CTMPs [17]. Approximate bisimilarity is also a central notion in this work. Not only does it provide a means for approximating a continuous-time Markov process, but also it is a technique to aggregate the state space. Approximate bisimilarity for CTMPs is also one topic that we are working on now.

Acknowledgments

We are grateful to Professor Ernst-Erich Doberkat for helpful suggestions on an earlier draft of this paper. We appreciate and want to thank the referees for invaluable comments which helped improve the paper dramatically.

[5] We want to thank Prof. E.-E Doberkat for this suggestion.

Appendix A. Completeness

Proof of Lemma 2.12. We prove this by induction on the complexity of φ.

- Base case when $\varphi := \top$. Obvious.
- Boolean cases. The proof is straightforward.
- Assume that $\varphi := L_r\psi$.

$$\begin{aligned} w \in f^{-1}([[\varphi]]_{M_2}) &\Leftrightarrow f(w) \in [[\varphi]]_{M_2} \\ &\Leftrightarrow M_2, f(w) \models L_r\psi \\ &\Leftrightarrow T_2(f(w))([[\psi]]_{M_2}) \geq r \\ &\Leftrightarrow T_1(w)(f^{-1}([[\psi]]_{M_2})) \geq r \\ &\Leftrightarrow T_1(w)([[\psi]]_{M_1}) \geq r \\ &\Leftrightarrow w \in [[L_r\psi]]_{M_1}. \end{aligned}$$

The fourth equivalence is based on the second condition in the definition of zigzag morphism and the fifth is according to the induction hypothesis. □

Proof of Lemma 2.13. We prove the lemma by induction on the complexity of φ.

- Base case $\varphi := \top$. Obvious.
- Boolean cases: the proof is straightforward.
- $\varphi := L_r\psi$.

$$\begin{aligned} \varphi \in d(w) &\Leftrightarrow M_1, w \models L_r\psi \\ &\Leftrightarrow T_1(w)([[\psi]]_{M_1}) \geq r \\ &\Leftrightarrow T_1(w)(f^{-1}([[\psi]]_{M_2})) \geq r \\ &\Leftrightarrow T_2(f(w))([[\psi]]_{M_2}) \geq r \\ &\Leftrightarrow L_r\psi \in d(f(w)). \end{aligned}$$

The third equivalence is based on Lemma 2.12 and the fourth is on Part (2) in the definition of zigzag morphism. □

In the following, we focus on the proof of the completeness. Whenever no confusion arises, we drop subscripts Σ_p in $\vdash_{\Sigma_p}$ and simply write $\vdash$.

Theorem A.1. *The following principles are provable in Σ_p:*

1. *If $\vdash \varphi \to \psi$, then $\vdash L_r\varphi \to L_r\psi$;*
2. *$\vdash L_r\varphi \to L_s\varphi$ if $r \geq s$;*

Proof. We reason inside Σ_p.

1. $\vdash \varphi \to \psi$ (Assumption)
   $\vdash \varphi \leftrightarrow \varphi \wedge \psi$ (A0)
   $\vdash L_r\varphi \leftrightarrow L_r(\varphi \wedge \psi)$ (DIS)
   $\vdash L_r(\varphi \wedge \psi) \wedge L_0(\neg\varphi \wedge \psi) \to L_r(\psi)$ (A3)
   $\vdash L_0(\neg\varphi \wedge \psi)$ (A1)
   $\vdash L_r(\varphi \wedge \psi) \to L_r(\psi)$ (A0)
   $\vdash L_r\varphi \to L_r\psi$ (A0)
2. If $r = t$, it is trivially true. Assume that $r > t$.
   $\vdash \neg L_t(\varphi \wedge \varphi) \wedge \neg L_{r-t}(\varphi \wedge \neg\varphi) \to \neg L_r\varphi$ (A4)
   $\vdash \neg L_{r-t}(\varphi \wedge \neg\varphi)$ (A2)
   $\vdash \neg L_t\varphi \to \neg L_r\varphi$ (DIS and A0)
   $\vdash L_r\varphi \to L_t\varphi$ (A0). □

Lemma A.2. *The following two propositions hold:*

1. *If $\vdash \neg(\varphi \wedge \psi)$, then $\vdash L_r\varphi \wedge L_s\psi \to L_{r+s}(\varphi \vee \psi)$, for $r + s \leq 1$;*
2. *If $\vdash \neg(\varphi \wedge \psi)$, then $\vdash \neg L_r\varphi \wedge \neg L_s\psi \to \neg L_{r+s}(\varphi \vee \psi)$, for $r + s \leq 1$.*

Proof. Assume that $\vdash \neg(\varphi \wedge \psi)$. It follows that $\vdash \varphi \to \neg\psi$ and $\vdash \psi \to \neg\varphi$. Moreover, $(\varphi \vee \psi) \wedge \varphi \leftrightarrow \varphi$, $(\varphi \vee \psi) \wedge \psi \leftrightarrow \psi$ and $(\varphi \vee \psi) \wedge \psi \to (\varphi \vee \psi) \wedge \neg\varphi$. It follows from (A3) that

$$\vdash L_r((\varphi \vee \psi) \wedge \varphi) \wedge L_s((\varphi \vee \psi) \wedge \neg\varphi) \to L_{r+s}(\varphi \vee \psi).$$

From (DIS) and (A0), we know

$$\vdash L_r\varphi \wedge L_s\psi \to L_r((\varphi \vee \psi) \wedge \varphi) \wedge L_s((\varphi \vee \psi) \wedge \neg\varphi).$$

The above two imply that:

$$\vdash L_r\varphi \wedge L_s\psi \to L_{r+s}(\varphi \vee \psi).$$

The proof of the second part is similar to that of the first part. □

In such a standard logic Σ as first order logic or normal modal logic, the deducibility relation $\vdash$ is defined only on propositional calculus:

$$\Gamma \vdash_{\Sigma} \varphi \text{ iff } \varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_n \to \varphi \text{ is a theorem of } \Sigma \text{ for some } \varphi_1, \varphi_2, \ldots, \varphi_n \in \Gamma.$$

So it holds for Σ that strong completeness = weak completeness + compactness. However, this equivalence is not true for *infinitary* logics like Σ_p whose deducibility relation is defined through such an additional infinitary rule as (ARCH). It is easy to see that compactness fails for Σ_p.

However, we show in the following that the above deductive system is indeed *strongly complete* with respect to the class of probability models.

Lemma A.3. 1. *If $\varphi \in \Gamma$, then $\Gamma \vdash \varphi$;*
2. *(Monotonicity) If $\Gamma \vdash \varphi$ and $\Gamma \subset \Delta$, then $\Delta \vdash \varphi$;*
3. *(Detachment) If $\Gamma \vdash \varphi$ and $\Gamma \vdash \varphi \to \psi$, then $\Gamma \vdash \psi$;*
4. *$\Gamma \vdash \varphi \to \psi$ implies $\Gamma, \varphi \vdash \psi$;*
5. *(Implication Rule) If $\Gamma \vdash \varphi$, then $\{\psi \to \gamma : \gamma \in \Gamma\} \vdash \psi \to \varphi$.*

Proof. All these are straightforward from the definition. See Chapter 9 in [26]. □

Definition A.4. A set Γ of formulas is

- *negation complete* if for every $\varphi \in \Phi_L$, $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$;
- *Σ_p-inconsistent* if $\Gamma \vdash_{\Sigma_p} \bot$, and *$\Sigma_p$-consistent* otherwise;
- *finitely Σ_p-consistent* if any finite subset of Γ is Σ_p-consistent;
- *maximally finitely Σ_p-consistent* if finitely Σ_p-consistent but no proper extension is finitely Σ_p-consistent;
- *maximally Σ_p-consistent* if Σ_p-consistent but no proper extension is Σ_p-consistent;
- *maximal* if it is negation complete and Σ_p-consistent. ◁

In the following sections, we drop Σ_p in front of consistency and just write consistency short for Σ_p-consistency.

Lemma A.5. *Γ is a set of formulas and φ is a formula.*

1. *If $\Gamma \vdash \varphi$ and $\Gamma \vdash \neg\varphi$, then Γ is inconsistent;*
2. *If Γ is finitely consistent, then so is one of $\Gamma \cup \{\varphi\}$ and $\Gamma \cup \{\neg\varphi\}$ for each φ.*
3. *If Γ is negation complete and finitely consistent, then it is closed under detachment.*
4. *Γ is maximally finitely consistent iff it is negation compete and finitely consistent.*
5. *Γ is maximal iff it is maximally consistent.*

Proof. The detailed proof is referred to Lemma 4.13 in [27]. □

Corollary A.6. *Assume that Γ is maximally consistent. Then*

1. *if $\Sigma \subseteq \Gamma$ and $\Sigma \vdash \varphi$, then $\varphi \in \Gamma$;*
2. *if $\vdash \varphi$, then $\varphi \in \Gamma$.*

Proof. This proposition follows directly from the above lemma. □

Corollary A.7. *For any formula φ, φ is consistent if and only if $\neg\varphi$ is not a theorem of Σ_s.*

Proof. Assume that φ is consistent.

$$\varphi \nvdash \bot \Leftrightarrow \nvdash \neg\varphi. \quad \square$$

This corollary also tells us that, for a *finite* set of formulas, the definition of its Σ_p-consistency is the same as that of consistency in normal modal logics or first order logic [8].

Definition A.8. A probability logic Σ_p is *Lindenbaum* if every Σ_p-consistent set of formulas has a maximal Σ_p-consistent extension. ◁

It is easy to the check that the consequence relation $\models$ is Lindenbaum.

Theorem A.9. *If Σ_p^i is a Lindenbaum probability logic for $i \in I$ where I is an index set and $\Sigma_p^\infty = \bigcap_i \Sigma_p^i$, then Σ_p^∞ is also a Lindenbaum probability logic.*

Proof. Assume that Γ is consistent in Σ_p^∞, i.e., $(\Gamma, \perp) \notin \Sigma_p^\infty$. This implies that $(\Gamma, \perp) \notin \Sigma_p^j$ for some $j \in I$. Since Σ_p^j is Lindenbaum, there is a maximally Σ_p^j-consistent set $\Gamma_0 \supseteq \Gamma$. But as $\Sigma_p^\infty \subseteq \Sigma_p^j$, Γ_0 is also Σ_p^∞-consistent. Of course, it is negation-complete, and hence is Σ_p^∞-maximal. $\square$

Definition A.10. Probability logic Σ_s is the least Lindenbaum probability logic. $\lhd$

In the following, we will show that a set of formulas Γ is maximally Σ_s-consistent iff it is a satisfied theory, i.e., the set of formulas which are satisfied at some point of a Markov process or probability model. In short, we will provide a proof of the completeness of our deductive system Σ_s with respect to the class of Markov processes. In the rest of the paper, consistency means Σ_s-consistency whenever no confusion arises.

Definition A.11. The canonical Markov process M^C is defined as follows:

- $\Omega_{\mathcal{L}_0} = \{\Gamma \subseteq \mathcal{L}_0 : \Gamma$ is a maximally consistent set of formulas$\}$.
- $[\varphi] = \{w \in \Omega_{\mathcal{L}_0} : \varphi \in w\}$;
- $\mathcal{A}_{\mathcal{L}_0}$ is the σ-algebra generated by $\{[\varphi] : \varphi \in \mathcal{L}_0\}$;
- The canonical measurable space is defined as $\langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0} \rangle$;
- define $T_{\mathcal{L}_0}(w)([\varphi]) = \sup\{r \in [0, 1] \cap Q : L_r\varphi \in w\}$ and we know that such a defined $T_{\mathcal{L}_0}(w)$ on the canonical algebra $\mathcal{A}_{\mathcal{L}_0}$ uniquely determine the subprobability measure $T_{\mathcal{L}_0}(w)$ on the σ-algebra $\mathcal{A}_{\mathcal{L}_0}$. This unique extension is guaranteed by Theorem 2.2. $\lhd$

Lemma A.12. *For any formulas φ and ψ,*

- $[\varphi] \subseteq [\psi]$ *iff* $\vdash \varphi \rightarrow \psi$;
- $[\varphi] = [\psi]$ *iff* $\vdash \varphi \leftrightarrow \psi$.

Proof. This lemma follows directly from the Lindenbaum property through the standard argument in modal logic [8]. $\square$

It follows from this lemma that the above $T_{\mathcal{L}_0}$ is well-defined. Note that $T_{\mathcal{L}_0}$ is total. Since $L_1\top$ is not necessarily a theorem in Σ_S, $T_{\mathcal{L}_0}(\top)$ might not be equal to 1. It remains to show that $T_{\mathcal{L}_0}(w)$ is a subprobability measure on $\mathcal{A}_{\mathcal{L}_0}$. First we show that the above defined $T_{\mathcal{L}_0}(w)$ is a finitely additive subprobability measure.

Lemma A.13. *For $A, B \in \mathcal{A}_{\mathcal{L}_0}$, $w \in \Omega_{\mathcal{L}_0}$, if $A \cap B = \emptyset$, then $T_{\mathcal{L}_0}(w)(A) + T^C(w)(B) = T_{\mathcal{L}_0}(w)(A \cup B)$.*

Proof. It is easy to see that there are formulas φ_1, φ_2 such that $A = [\varphi_1]$, $B = [\varphi_2]$ and $\vdash \varphi_1 \rightarrow \neg\varphi_2$. Let α_1, α_2 and α_+ denote $T_{\mathcal{L}_0}(w)([\varphi_1])$, $T_{\mathcal{L}_0}(w)([\varphi_2])$ and $T_{\mathcal{L}_0}(w)([\varphi_1 \vee \varphi_2])$, respectively. So we only need to show that $\alpha_1 + \alpha_2 = \alpha_+$.

Suppose that $\alpha_1 + \alpha_2 < \alpha_+$. Then there are $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $(\alpha_1 + \epsilon_1) + (\alpha_2 + \epsilon_2) < \alpha_+$, $\alpha_1 + \epsilon_1 \in Q$ and $\alpha_2 + \epsilon_2 \in Q$. Let $\alpha_1' := \alpha_1 + \epsilon_1$ and $\alpha_2' := \alpha_2 + \epsilon_2$. It follows that $L_{\alpha_1'}\varphi_1 \notin w$ and hence $\neg L_{\alpha_1'}\varphi_1 \in w$. Similarly, $\neg L_{\alpha_2'}\varphi_2 \in w$. By (A4) (actually by Lemma A.2(2) and Corollary A.6), we know that $\neg L_{\alpha_1'+\alpha_2'}(\varphi_1 \vee \varphi_2) \in w$. But this is impossible because $\alpha_1' + \alpha_2' < \alpha_+$ and hence $L_{\alpha_1'+\alpha_2'}(\varphi_1 \vee \varphi_2) \in w$.

The following argument is dual to the above one. Suppose that $\alpha_1 + \alpha_2 > \alpha_+$. Then there are two $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $(\alpha_1 - \epsilon_1) + (\alpha_2 - \epsilon_2) > \alpha_+$, $\alpha_1 - \epsilon_1 \in Q$ and $\alpha_2 - \epsilon_2 \in Q$. Let $\alpha_1'' := \alpha_1 - \epsilon_1$ and $\alpha_2'' := \alpha_2 - \epsilon_2$. It follows that $M_{\alpha_1''}\varphi_1 \notin w$ and hence $L_{\alpha_1''}\varphi_1 \in w$. Similarly, $L_{\alpha_2''}\varphi_2 \in w$. Since $\vdash \varphi_1 \rightarrow \neg\varphi_2$, by (A3), we know that $L_{\alpha_1''+\alpha_2''}(\varphi_1 \vee \varphi_2) \in w$. But this is impossible because $\alpha_1'' + \alpha_2'' > \alpha_+$ and hence $L_{\alpha_1''+\alpha_2''}(\varphi_1 \vee \varphi_2) \notin w$. $\square$

Theorem A.14. *The above defined $T_{\mathcal{L}_0}(w)$ is a subprobability measure on $\mathcal{A}_{\mathcal{L}_0}$.*

Proof. According to Theorem 2.1, it suffices to show that if $\{[\varphi_n] : n \in \mathbf{N}\}$ is a non-increasing sequence of sets in $\mathcal{A}_{\mathcal{L}_0}$ whose intersection is empty, i.e., $[\varphi_1] \supseteq [\varphi_2] \supseteq \cdots \supseteq [\varphi_n] \supseteq \cdots$ and $\bigcap_n[\varphi_n] = \emptyset$, then $\lim_{n\to\infty} T^C(w)([\varphi_n]) = 0$ for all $w \in \Omega_{\mathcal{L}_0}$.

Note that finite additivity of $T_{\mathcal{L}_0}(w)$ implies its monotonicity. That is to say, $T_{\mathcal{L}_0}(w)([\varphi_1]) \geq T_{\mathcal{L}_0}(w)([\varphi_2]) \geq \cdots \geq T_{\mathcal{L}_0}(w)([\varphi_n]) \geq \cdots$. Now we show that $\lim_{n\to\infty} T_{\mathcal{L}_0}(w)([\varphi_n]) = 0$ by contradiction.

Suppose that the limit was positive, then there is a rational p such that $T_{\mathcal{L}_0}(w)([\varphi_n]) > p > 0$ for all n. Consider the set $\Gamma := \{\varphi_n : n \in \mathbf{N}\}$. Recall that $\bigwedge_\omega \Gamma$ is the set of all conjunctions of finite subsets of Γ. For any $\psi \in \bigwedge_\omega \Gamma$, one has that $[\psi] = [\varphi_m]$ for some m. According to the assumption, we know that $T(w)([\psi]) > p$, this implies that $L_p\psi \in w$. This also means that the set $L_p(\bigwedge_\omega \Gamma) = \{L_p\psi : \psi \in \bigwedge_\omega \Gamma\} \subseteq w$. From **A**2, we know that $\vdash \neg L_p\perp$ (p is positive) and hence $\neg L_p\perp \in w$. Since w is a maximally consistent set of formulas, $L_p(\bigwedge_\omega \Gamma)$ is consistent. Therefore $L_p(\bigwedge_\omega \Gamma) \nvdash L_p\perp$. It follows from the **rule (CAR)** that $\Gamma \nvdash \perp$. That is to say Γ is consistent.

According to the above **Lindenbaum property**, there is a maximally consistent set w' of formulas such that $\Gamma \subseteq w'$ and hence $\varphi_n \in w'$ for all n. Equivalently, $w' \in \bigcap_n[\varphi_n]$. This also means that $\bigcap_n[\varphi_n] \neq \emptyset$, which contradicts our assumption that $\bigcap_n[\varphi_n] = \emptyset$.

So we conclude that $\lim_{n\to\infty} T_{\mathcal{L}_0}(w)([\varphi_n]) = 0$. Theorem 2.1 guarantees that $T_{\mathcal{L}_0}(w)$ is also countably additive on the σ-algebra $\sigma(\mathcal{A}_0^C)$. Therefore each $T_{\mathcal{L}_0}(w)$ defined on the canonical model is a well-defined probability measure. $\square$

This theorem is the only place where we need the Countable Additivity Rule.

Corollary A.15. *The above defined canonical Markov process $\langle \Omega_{\mathcal{L}_0}, \mathcal{A}_{\mathcal{L}_0}, T_{\mathcal{L}_0} \rangle$ is a Markov process.*

Proof. Here we must note that Theorem 2.3 in Section 2 plays an essential role in the proof that $T_{\mathcal{L}_0}$ is a transition subprobability function. □

In Theorem 2.17, we will show that the canonical Markov process is actually on a *final* coalgebra.

Lemma A.16 (*Truth Lemma*)**.** *For any $w \in \Omega_{\mathcal{L}_0}$ and any $\varphi \in L_0$, $M_{\mathcal{L}_0}, w \models \varphi$ if and only if $\varphi \in w$, i.e. $[[\varphi]]_{M_{\mathcal{L}_0}} = [\varphi]$.*

Proof. We prove by induction on the complexity of φ. Here we only deal with the non-trivial case.

$$\begin{aligned} L_r\varphi \in w \; &\Leftrightarrow r \le T_{\mathcal{L}_0}(w)([\varphi]) \\ &\Leftrightarrow M_{\mathcal{L}_0}, w \models L_r\varphi \end{aligned}$$

We use the rule (ARCH) in the proof of the right-to-left direction of the first equivalence. The last step is based on the induction hypothesis that $[\varphi] = [[\varphi]]_{M_{\mathcal{L}_0}}$. □

Proof of Theorem 2.16. It suffices to show that any Σ_s-consistent set of formulas is satisfiable in a probability model. But this follows immediately from the Lindenbaum property and the Truth Lemma. □

Proof of Theorem 2.17. We prove that d satisfies the three conditions in the definition of zigzag morphism.

1. First we show that, for each $[\varphi] \in \mathcal{A}^0_{\mathcal{L}_0}$, $d^{-1}([\varphi]) = [[\varphi]]_M \in \mathcal{A}$.

$$\begin{aligned} w \in d^{-1}([\varphi]) \; &\Leftrightarrow d(w) \in [\varphi] \\ &\Leftrightarrow \varphi \in d(w) \\ &\Leftrightarrow M, w \models \varphi \\ &\Leftrightarrow w \in [[\varphi]]_M. \end{aligned}$$

2. For any $w \in \Omega$, $[\varphi] \in \mathcal{A}^0_{\mathcal{L}_0}$,

$$T_{\mathcal{L}_0}(d(w))([\varphi]) \quad = \tag{1}$$
$$sup\{r \in Q \cap [0, 1] : L_r\varphi \in d(w)\} \quad = \tag{2}$$
$$\begin{aligned} sup\{r \in Q \cap [0, 1] : M, w \models L_r\varphi\} &= T(w)([[\varphi]]_M) \\ &= T(w)(d^{-1}([\varphi])) \end{aligned}$$

The last equality came from the first part.

Indeed d is a zigzag morphism from M to $M_{\mathcal{L}_0}$. The uniqueness follows from Lemma 2.13. So indeed $M_{\mathcal{L}_0}$ is final in the class of Markov processes. □

Appendix B. Isomorphism between final coalgebras

In this section, the terminology is from [36] and we fix the functor $T := \Delta(Id + \mathbf{1})$, which is a measurable polynomial functor.

B.1. Syntax and semantics

$Ing(\Delta(Id + \mathbf{1})) = \{Id, Id + \mathbf{1}, \mathbf{1}, \Delta(Id + \mathbf{1})\}$. We define a language $\mathcal{L}(T)$ as follows: for any $S \in Ing(T)$

$$true_S : S \qquad \emptyset : \mathbf{1} \qquad \{1\} : \mathbf{1}$$
$$\frac{\varphi : S, \psi : S}{\varphi \wedge \psi : S} \qquad \frac{\varphi : Id}{inl_{Id+\mathbf{1}}\varphi : Id + \mathbf{1}} \qquad \frac{\psi : \mathbf{1}}{inr_{Id+\mathbf{1}}\psi : Id + \mathbf{1}}$$
$$\frac{\varphi : Id + \mathbf{1}, p \in [0, 1]}{\beta^p\varphi : \Delta(Id + \mathbf{1})} \qquad \frac{\varphi : \Delta(Id + \mathbf{1})}{[next]\varphi : Id}$$

Let $c : X \to TX$ be a coalgebra of T. The semantics of $\mathcal{L}(T)$ is defined as follows:

$$[[true]]^c_S = SX \qquad [[\emptyset]]^c_{\mathbf{1}} = \emptyset \qquad [[\{1\}]]^c_{\mathbf{1}} = \{1\}$$
$$[[\varphi \wedge \psi]]^c_S = [[\varphi]]^c_S \cap [[\psi]]^c_S \qquad [[inl\varphi]]^c_{Id+\mathbf{1}} = inl([[\varphi]]^c_{Id}) \qquad [[inr\varphi]]^c_{Id+\mathbf{1}} = inr([[\varphi]]^c_{\mathbf{1}})$$
$$[[\beta^p\varphi]]^c_{\Delta(Id+\mathbf{1})} = \beta^p([[\varphi]]^c_{Id+\mathbf{1}}) \qquad [[[next]\varphi]]^c_{Id} = c^{-1}([[\varphi]]^c_T)$$

where $\beta^p([[\varphi]]^c_{Id+\mathbf{1}}) = \{\mu \in \Delta(Id + \mathbf{1}) : \mu([[\varphi]]^c_{Id+1}) \ge p\}$.

B.2. Final coalgebra for $\Delta(Id + \mathbf{1})$

Definition B.17. For each coalgebra $c : x \to TX$ and each $x \in SX$ where $S \in Ing(T)$, we define

$$d^c_S(x) := \{\varphi : S | x \in [[\varphi]]^c_S\}.$$

Each such set $d^c_S(x)$ is called a *satisfied theory*. $\lhd$

Definition B.18. We define the canonical sets S^* for $S \in Ing(T)$ by

$$S^* = \{d^c_S(x) : x \in SX \text{ for some coalgebra } c : X \to TX\}. \quad \lhd$$

So Id^* is the "maximally consistent set" of formulas $[next]\varphi : Id$ or $true_{Id} : Id$. The following two propositions are from Lemma 5.1 and Theorem 6.4 in [36].

Lemma B.19. *There is a family of measurable maps $r_S : S^* \to S(Id^*)$ indexed by the ingredients of T such that, for all coalgebras $c : X \to TX$, the diagram below commutes:*

$$\begin{tikzcd} S^* \arrow[r, "r_S"'] & S(Id^*) \\ SX \arrow[u, "d^c_S"'] \arrow[ur, "Sd^c_{Id}"'] & \end{tikzcd}$$

Theorem B.20. *Let $c^* : Id^* \to T(Id^*)$ be*

$$r_T \circ [next]^{-1} : Id^* \to T^* \to T(Id^*)$$

$c^* := r_T \circ [next]^{-1}$ is the final coalgebra for T.

B.3. Translation

In this section we define an inter-translation between formulas with and without sort *Id*. From the syntax of $\mathcal{L}(T)$ we know that

- each formula of sort *Id* is $true_{Id}$ or is of the form $\bigwedge_i [next]\varphi_i$ where
- each φ_i is of the form $\bigwedge_i \beta^{p_i}\psi_i$ (here $true_{\Delta Id}$ is taken to be $\beta^1(true_{Id})$) where
- each ψ_i is of the form $\bigwedge_k inl_{Id+\mathbf{1}}\theta_k$ or of the form $\bigwedge_k inr_{Id+\mathbf{1}}\theta_k$ where
- each θ_k is of the sort *Id*.

So each formula of sort *Id* is of the following form

$$true_{Id} \wedge \bigwedge_{i \in I} [next] \left(\bigwedge_{j \in J_i} \beta^{p_i}\psi_j \right)$$

where ψ_j is of the form $(\bigwedge_{k \in K_j} inl_{Id+\mathbf{1}}\theta_k) \wedge (\bigwedge_{k' \in K'_j} inr_{Id+\mathbf{1}}\theta_{k'})$. We stipulate that if $I = \emptyset$ or $J = \emptyset$, then the conjunction is equivalent to $true_{Id}$.

Recall that the simple probabilistic modal logic $\mathcal{L}_0$ is defined as follows:

$$\varphi := \top \mid \varphi \wedge \varphi \mid L_p\varphi.$$

Definition B.21. Let $c : X \to T(X)$ be a coalgebra for T where $\mathbf{X} = \langle X, \mathcal{A} \rangle$ is a measurable space. The *satisfaction relation* on $X \times \mathcal{L}_0$ is defined inductively as follows:

- $(X, c), x \models \top$ for all $x \in X$;
- $(X, c), x \models \varphi_1 \wedge \varphi_2$ iff $(X, c), x \models \varphi_1$ and $(X, c), x \models \varphi_2$;
- $(X, c), x \models L_p\varphi$ iff $c(x)(inl([[\varphi]]^c)) \geq p$ where $[[\varphi]]^c = \{x \in X : (X, c), x \models \varphi\}$.

$d^c(x)$ is the set of formulas in $\mathcal{L}_0$ that are satisfied at x in (X, c) and is called the *satisfied theory* of formulas at x. $\lhd$

For each $x \in X$ and $A \in \mathcal{A}$, define $T(x, A) := c(x)(inl(A))$. It is easy to see that such defined T is a transition subprobability function and $\langle X, \mathcal{A}, T \rangle$ is a Markov process.

Definition B.22. A *translation* from $^\circ : \mathcal{L}(T) \to \mathcal{L}_0$ is defined inductively as follows:

- $(true_{Id} : Id)^\circ = \top$;
- $(\varphi \wedge \psi : Id)^\circ = (\varphi)^\circ \wedge (\varphi)^\circ$;
- If ψ_j is of the form $(\bigwedge_{k \in K_j} inl_{Id+\mathbf{1}}\theta_k) \wedge (\bigwedge_{k' \in K'_j} inr_{Id+\mathbf{1}}\theta_{k'})$, then $([next](\bigwedge_{j \in J} \beta^{p_i}\psi_j) : Id)^\circ = \bigwedge_{j \in J} L_{p_i}((\bigwedge_{k \in K_j} \theta^\circ_k)$. $\lhd$

The third clause in the translation is the most important one. The intuition behind this clause is formalized in the following proposition. For a set of formulas $\Gamma \subseteq \mathcal{L}(T)$ of sort *Id*, $\Gamma^\circ := \{\varphi^\circ : \varphi \in \Gamma\}(\subseteq \mathcal{L}_0)$. We can easily prove by induction the following proposition.

Theorem B.23. *Let $c : X \to T(X)$ be a coalgebra. For any formula $\varphi : Id$ in $\mathcal{L}(T)$,*

$$[[\varphi]]^c_{Id} = [[\varphi^\circ]]^c$$

Corollary B.24. *Let $c : X \to T(X)$ be a coalgebra. The above two description maps d^c_{Id} and d^c are equivalent in the sense that, for all $x \in X$,*

$$(d^c_{Id}(x))^\circ = d^c(x).$$

Goldblatt [27] provided a deduction system $\vdash_{\mathcal{L}(T)}$ for $\mathcal{L}(T)$ and in Section 2 we adapted it to be a complete deductive system $\vdash_{\mathcal{L}_0}$ for $\mathcal{L}_0$. This theorem follows from Theorem 5.17 in [27].

Theorem B.25. *Let Γ be a set of formulas in $\mathcal{L}(T)$ of sort Id. Γ is maximal $\vdash_{\mathcal{L}(T)}$-consistent iff it is $d^c_{Id}(x)$ for some state x in some coalgebra $c : X \to T(X)$.*

B.4. Isomorphism between final coalgebras of $\Delta(Id + \mathbf{1})$

Recall that Id^* is the set of all satisfied theories of formulas of sort *Id* and is equipped with the σ-algebra $\mathcal{A}_{ID}$ that is generated by the sets $|\varphi|_{Id}$ where $|\varphi|_{Id} := \{\Gamma \in Id^* : \varphi \in \Gamma\}$. $(Id^*)^\circ = \{(d^c_{Id}(x))^\circ : x$ is a state in some coalgebra $c : X \to \Delta(X)\}$. The proposition below follows directly from Theorem B.23.

Corollary B.26. *$(Id^*)^\circ$ is the class of* all *maximally $\vdash_{\mathcal{L}_0}$-consistent sets of formulas in $\mathcal{L}_0$. The translation $^\circ$ is one-to-one on Id^* in the sense that, if $d^c_{Id}(x) \neq d^{c'}_{Id}(x')$ for states x in $c : X \to T(X)$ and x' in $c' : X' \to T(X')$, then $(d^c_{Id}(x))^\circ \neq (d^{c'}_{Id}(x'))^\circ$, namely, $d^c(x) \neq d^{c'}(x')$.*

First recall some notions from Appendix A. $\Omega_{\mathcal{L}_0}$ is the class of all maximally $\vdash_{\mathcal{L}_0}$-consistent sets of formulas in $\mathcal{L}_0$. So it is the same as $(Id^*)^\circ$. We equip $\Omega_{\mathcal{L}_0}$ with the σ-algebra $\mathcal{A}_{\mathcal{L}_0}$ generated by the sets $|\varphi| := \{\Gamma \in \Omega_{\mathcal{L}_0} : \varphi \in \Gamma\}$, i.e., maximally $\vdash_{\mathcal{L}_0}$-consistent sets of formulas in $\mathcal{L}_0$. Note that, for a formula $\varphi : Id$ in $\mathcal{L}(T)$, $(|\varphi|_{Id})^\circ = |\varphi^\circ|$. For any $s \in \Omega_{\mathcal{L}_0}$ and $|\varphi| \subseteq \Omega_{\mathcal{L}_0}$, define

$$T_{\mathcal{L}_0}(s)(|\varphi|) = \{p \in [0, 1] : L_p\varphi \in s\}.$$

We have shown in Appendix A that $T_{\mathcal{L}_0}(s)$ is a finitely additive measure on the set of $|\varphi|$'s and it can be uniquely extended to a measure on $\mathcal{A}_{\mathcal{L}_0}$. We equip $\Delta(\Omega_{\mathcal{L}_0})$ with the σ-algebra $\mathcal{A}_{\mathcal{L}_0}$ generated by the following sets

$$\beta^p(|\varphi|) := \{\mu \in \Delta(\Omega_{\mathcal{L}_0}) : \mu(|\varphi|) \geq p\}.$$

Define $c_{\mathcal{L}_0}$ to be a function from $\Omega_{\mathcal{L}_0}$ to $T(\Omega_{\mathcal{L}_0})$ satisfying the condition $c_{\mathcal{L}_0}(s)(inl(|\varphi|)) = T_{\mathcal{L}_0}(s)(|\varphi|)$. It is easy to see that such defined $c_{\mathcal{L}_0} : \Omega_{\mathcal{L}_0} \to T(\Omega_{\mathcal{L}_0})$ is $\mathcal{A}_{\mathcal{L}_0}$-measurable and hence is a coalgebra for T.

Lemma B.27. *Let $c : X \to T(X)$. For each formula $\varphi \in \mathcal{L}_0$,*

$$c(x)(inl([[\varphi]]^c)) = c_{\mathcal{L}_0}(d^c(x))(inl(|\varphi|)).$$

Proof. Both sides are equal to $\sup\{p : (X, c), x \models L_p\varphi\}$. The proof of this lemma is essentially similar to that of Theorem 2.17. $\square$

Theorem B.28 (*Main Theorem*)**.** *The following diagram commutes:*

$$\begin{tikzcd} Id^* \arrow[r, "{[next]^{-1}}"] \arrow[d, "\circ"] & T^* \arrow[r, "r_T"] & T(Id^*) \arrow[d, "\Delta\cdot^\circ"] \\ \Omega_{\mathcal{L}} \arrow[rr, "c_{\mathcal{L}_0}"] & & T(\Omega_{\mathcal{L}_0}) \end{tikzcd}$$

That is to say, $^\circ : Id^* \to \Omega_{\mathcal{L}_0}$ is a coalgebra homomorphism. Moreover, the inverse of the translation $^\circ$ is also coalgebra homomorphism. So, the following two coalgebras are isomorphic:

- $c^* : Id^* \to T(Id^*)$;
- $c_{\mathcal{L}_0} : \Omega_{\mathcal{L}_0} \to T(\Omega_{\mathcal{L}_0})$.

Proof. Let $c : X \to T(X)$ be a coalgebra. $c_{Id+\mathbf{1}} : X \to X + \mathbf{1}$ is defined by $c_{Id+\mathbf{1}}(x) = inl(x)$. It is easy to see that there is a unique measurable function $c_\Delta : X + \mathbf{1} \to \Delta(X + \mathbf{1})(= T(X))$ such that $c = c_\Delta c_{Id+\mathbf{1}}$.

Given a formula $\psi \in \mathcal{L}_0$, there is a formula $\varphi : Id$ in $\mathcal{L}(T)$ such that $\varphi^\circ = \psi$. Note that $(|\varphi|_{Id})^\circ = |\varphi^\circ|$. Moreover, from Theorem B.23, we know that the inverse image of $|\varphi^\circ|$ under $(\_^\circ)^{-1}$ is $|\varphi|_{Id}$. It suffices to show the following equality:

$$c^*(d^c_{Id}(x))(inl(|\varphi|_{Id})) = c_{\mathcal{L}_0}((d^c_{Id}(x))^\circ)(inl(|\varphi^\circ|))$$

We reason as follows:

$$\begin{aligned}
c^*(d^c_{Id}(x))(inl(|\varphi|_{Id})) &= r_T \circ [next]^{-1}(d^c_{Id}(x))(inl(|\varphi|_{Id})) \quad \text{(Theorem B.20)}\\
&= r_T(d^c_T(c(x)))(inl(|\varphi|_{Id})) \quad \text{(Lemma 4.12 in [36])}\\
&= (Td^c_{Id})(c(x))(inl(|\varphi|_{Id})) \quad \text{(Lemma B.19)}\\
&= (\Delta(Id+\mathbf{1})d^c_{Id})(c_\Delta c_{Id+\mathbf{1}}(x))(inl(|\varphi|_{Id}))\\
&= c_\Delta(c_{Id+\mathbf{1}}(x))((Id+\mathbf{1})d^c_{Id})^{-1}(inl(|\varphi|_{Id}))) \quad \text{(Definition of } \Delta)\\
&= c(x)(inl([[\varphi]]^c_{Id}))\\
&= c(x)(inl([[\varphi^\circ]]^c)) \quad \text{(Theorem B.23)}\\
&= c_{\mathcal{L}_0}(d^c(x))(inl(|\varphi^\circ|)) \quad \text{(Theorem B.27)}. \quad \square
\end{aligned}$$

So the final coalgebra for the functor $\Delta(Id + \mathbf{1})$ using formulas in $\mathcal{L}_0$ as in Section 2 is isomorphic to the one for the functor $\Delta(Id + \mathbf{1})$ from [36].

References

[1] L. de Alfaro, R. Majumdar, V. Raman, M. Stoelinga, Game refinement relations and metrics, Log. Methods Comput. Sci. 4 (3) (2008).
[2] A. Asperti, G. Longo, Categories, Types and Structures. Foundation of Computing, first edition, MIT Press, 1991.
[3] R. Aumann, Interactive epistemology: probability, Internat. J. Game Theory 28 (1999) 301–314.
[4] R. Aumann, A. Heifetz, Incomplete information, in: R. Aumann, S. Hart (Eds.), Handbook of Game Theory, in: Handbook Series in Economics, 2002, pp. 1665–1686.
[5] A. Aziz, K. Sanwal, V. Singhal, R. Brayton, Model-checking continous-time markov chains, ACM Trans. Comput. Log. 1 (1) (2000) 162–170.
[6] C. Baier, B. Haverkort, H. Hermanns, J. Katoen, Model-checking algorithms for continuous-time markov chains, IEEE Trans. Software Eng. 29 (6) (2003) 524–541.
[7] P. Billingsley, Measure and Probability, third edition, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc, 1995.
[8] P. Blackburn, M. de Rijke, Y. Venema, Modal Logic, in: Cambridge Tracts in Theoretical Computer Science, vol. 53, Oxford University Press, 2000.
[9] P. Chaput, V. Danos, P. Panangaden, G.D. Plotkin, Approximating markov processes by averaging, in: ICALP (2), 2009, pp. 127–138.
[10] V. Danos, J. Desharnais, P. Panangaden, Labelled markov processes: stronger and faster approximations, Electron. Notes Theor. Comput. Sci. 87 (2004) 157–203.
[11] V. Danos, Josee Desharnais, François Laviolette, Prakash Panangaden, Bisimulation and cocongruence for probabilistic systems, Inform. and Comput. 204 (4) (2006) 503–523.
[12] P. de Vink, J. Rutten, Bisimulation for probabilistic transition systems: a coalgebraic approach, Theoret. Comput. Sci. 221 (1–2) (1999) 271–293.
[13] J. Desharnais, A. Edalat, P. Panangaden, A logical characterization of bisimulation for labeled markov processes, in: Proceedings of the 13th IEEE Symposium on Logic in Computer Science, 1998, pp. 478–489.
[14] J. Desharnais, A. Edalat, P. Panangaden, Bisimulation for labeled markov processes, Inform. and Comput. 179 (2002) 163–193.
[15] J. Desharnais, V. Gupta, R. Jagadeesan, P. Panangaden, Approximating labeled markov processes, Inform. and Comput. 184 (2003) 160–200.
[16] J. Desharnais, V. Gupta, R. Jagadeesan, P. Panangaden, Metrics for labeled markov processes, Theoret. Comput. Sci. 318 (2004) 323–353.
[17] J. Desharnais, P. Panangaden, Continuous stochastic logic characterizes bisimulation of continuous-time markov processes, J. Log. Algebr. Program. 56 (1–2) (2003) 99–115.
[18] E. Doberkat, C. Schubert, Coalgebraic logic over general measurable spaces — a survey, Math. Structures Comput. Sci. 21 (2) (2011) 175–234.
[19] E.E. Doberkat, Topological completeness in an ideal model for recursive polymorphic types, SIAM J. Computing 18 (1989) 977–991.
[20] E.E. Doberkat, Semi-pullbacks for stochastic relations over analytic spaces, Math. Structures Comput. Sci. 115 (2005) 647–670.
[21] E.E. Doberkat, The Hennessy–Milner equivalence for continuous time stochastic logic with mu-operator, J. Appl. Log. 5 (3) (2006) 519–544.
[22] E.E. Doberkat, Stochastic relations: congruence, bisimulations and the Hennessy–Milner theorem, SIAM J. Computing 35 (3) (2006) 590–626.
[23] E.E. Doberkat, Stochastic Coalgebraic Logic, in: EATCS Mongraphs in Theoretical Computer Sciences, Springer, 2010.
[24] G. Edgar, Integral, Probability, and Fractal Measures, Springer, 1998.
[25] M. Giry, A categorical approach to probability theory, in: Categorical Aspects of Topology and Analysis, in: Lecture Notes in Mathematics, vol. 915, Springer, 1982, pp. 68–85.
[26] R. Goldblatt, Mathematics of Modality, in: CSLI Lecture Notes, vol. 43, CSLI Publications, Stanford, 1993.
[27] R. Goldblatt, Deduction systems for coalgebras over measurable spaces, J. Logic Comp. 20 (5) (2010) 1069–1100.
[28] P. Halmos, Measure Theory, in: Graduate Text in Mathematics, Van Nostrand, New York, 1950.
[29] A. Heifetz, P. Mongin, Probability logic for type spaces, Games Econom. Behav. 35 (2001) 31–53.
[30] A. Heifetz, D. Samet, Topology-free typology of beliefs, J. Econom. Theory 82 (1998) 324–341.
[31] M. Hennessy, R. Milner, Algebraic laws for nondeterminism and concurrency, J. ACM 32 (1) (1985) 137–161.
[32] B. Jacobs, Introduction to Coalgebra: Towards Mathematics of States and Observations, 2008. Book Draft.
[33] B. Jacobs, A. Sokolova, Exemplaric expressivity of modal logics, J. Logic Comput. 20 (5) (2010) 1041–1068.
[34] D. Kozen, A probabilistic pdl, J. Comput. System Sci. 30 (1985) 162–178.
[35] R. Milner, Communication and Concurrency, Prentice-Hall, NJ, 1989.
[36] L. Moss, I. Viglizzo, Final coalgebras for functors on measurable spaces, Inform. and Comput. 204 (2006) 610–636.
[37] J. Rutten, Universal coalgebra: a theory of systems, Theoret. Comput. Sci. 249 (1) (2000) 3–80.
[38] A. Sokolova, Probabilistic systems, coalgebraically: a survey, Theoret. Comput. Sci. 412 (38) (2011) 5095–5110.
[39] F. van Breugel, Claudio Hermida, Michael Makkai, James Worrell, Recursively defined metric spaces without contraction, Theoret. Comput. Sci. 380 (1–2) (2007) 143–163.
[40] F. van Breugel, J. Worrell, A behavioural pseudometric for probabilistic transition systems, Theoret. Comput. Sci. 331 (1) (2005) 115–142.

[41] F. van Breugel, James Worrell, Approximating and computing behavioural distances in probabilistic transition systems, Theoret. Comput. Sci. 360 (1–3) (2006) 373–385.
[42] M. Ying, pi-calculus with noisy channels, Acta Inform. 41 (9) (2005) 525–593.
[43] M. Ying, M. Wirsing, Approximate bisimilarity, in: AMAST, 2000, pp. 309–322.
[44] C. Zhou, Complete deductive systems for probability logic with application in Harsanyi type spaces, Ph.D. Thesis, Indiana University, Bloomington, 2007.
[45] C. Zhou, A complete deductive system for probability logic, J. Logic Comput. 19 (6) (2009) 1427–1454.
[46] C. Zhou, Probability logics for finitely additive beliefs, J. Log. Lang. Inf. 19 (3) (2010) 247–282.